Slide 1

Focus projection between theory and evidence

Towards using corpora for research linking syntax, prosody, and information structure

Detmar Meurers
University of Tübingen

joint work with Kordula De Kuthy

UFR d'Études Anglophones
Université Paris Diderot, Paris 7
March 11, 2010

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Slide 2

Motivation

▸ The nature of the integration of a sentence into discourse can provide an explanation 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.

▸ German and English are intonation languages:

   ▸ Prosody plays an important role in constraining the possible integration of a sentence into the discourse.

▸ Research relating syntax, information structure, and intonation predominantly theoretically driven

   ▸ partly contradictory empirical assumptions

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

   ⇒ A pilot study based on intonationally and syntactically annotated corpora.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Slide 3

Expressing information structure

▸ Languages differ with respect to how the information structure of an utterance is represented.

▸ Linguistic means of marking information structure include:
   ▸ word order
   ▸ morphology
   ▸ prosody

▸ English and German are so-called intonation languages
   ▸ Information structuring is signaled by the intonation (contour) of an utterance, including pitch accents.

   ▸ The absence or presence of an accent is an indicator of the discourse function of a constituent in a sentence.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Slide 4

Tones and Break Indices (ToBI)

▸ ToBI is a system for transcribing the intonation patterns and other aspects of the prosody of English utterances.

   ▸ Based on autosegmental-metrical approach to intonation. (cf. Beckman & Pierrehumbert 1986)

▸ The perceived intonation contour is described in terms of high (H) and low (L) targets in the local pitch range.

▸ ToBI transcription marks:

   ▸ 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 (with break values from 0–4):
      ▸ intermediate boundary (0–3): H⁻, L⁻
      ▸ full boundary (4): L% or H%

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

An example from the IMS Radionewscorpus

(1) *Das Gesetz tritt am kommenden Montag in Kraft.*
the law turns on the following Monday into power
*L*H* *!H*L?* *L%*

Das	Gesetz	tritt	am	kommenden	Montag	in	Kraft	-
das1	g@0 zEts1 (L*H)	tRIt1	am1	kOm1 m@n0 d@n0	mo:n1 (!H*L?) ta:k0	In1	kRaft1 (L%)	$.
ART	NN	VVFIN	APPRART	ADJA	NN	APPR	NN	0
1	1	1	1	1	1	1	4	
	L*H				!H*L?		L%	

Focus Projection Between Theory and Evidence
Detmar Meurers
University of Tübingen

Motivation
Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence
IMS Radionews Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice
Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep-use
Corpus exploration summary
Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

6 / 28

Discourse function of pitch accents

- ▸ Pitch accents can have different shapes allowing them to signal different functions in the discourse.
- ▸ The H* pitch accents are often assumed to signal focus, i.e., new material.
  - ▸ A more precise mapping of each possible pitch accent to its information structure role is subject of debate.
- ▸ Autosegmental-metrical approach to German intonation (Uhmann 1991; Féry 1993) assumes:
  - ▸ H* or H*+L represent focus accents
  - ▸ L* highlights background constituents
  - ▸ L*+H is a topic marker

Relating intonation and interpretation
Focus projection

- ▸ The word marked by a pitch accent and the extension of the focus are traditionally related by focus projection.

(2)
a. *What did John rent?* (narrow, NP focus)
John rented [[a BICYCLE]]$_F$.
b. *What did John do?* (wide, VP focus)
John [rented a BICYCLE]$_F$.
c. *What happened yesterday?* (wide, S focus)
[John rented a BICYCLE]$_F$.

Focus Projection Between Theory and Evidence
Detmar Meurers
University of Tübingen

Motivation
Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence
IMS Radionews Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice
Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
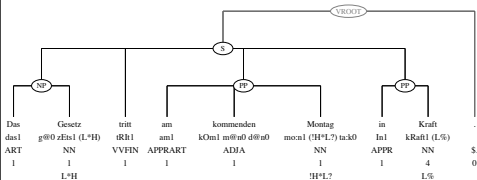Variation in downstep-use
Corpus exploration summary
Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

7 / 28

From pitch accent to projected focus
Traditional focus projection approach

Focus Projection Between Theory and Evidence
Detmar Meurers
University of Tübingen

Motivation
Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence
IMS Radionews Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice
Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep-use
Corpus exploration summary
Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

8 / 28

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
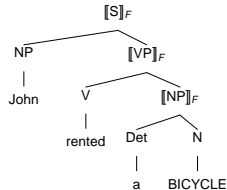
An open issue

Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

9/28

Constraining focus projection

► Focus projection rules determine the focus projection potential of a pitch accent depending on syntactic structure

► Strongest accent is generally constrained to fall on the last element (e.g., Nuclear Stress Rule of Chomsky & Halle 1968 for English; Jacobs 1988, p. 124 for German)

► Narrow focus examples exemplifying this:

   (3) *Who rented a bicycle yesterday?*
     ⟦*JOHN*⟧$_F$ *rented a bicycle.*

   (4) *Did John buy a bicycle?*
     John ⟦*RENTED*⟧$_F$ *a bicycle.*

These are not felicitous answers to questions such as:

   (5) a. *What did John rent?*
      b. *What did John do?*
      c. *What happened yesterday?*

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection

An open issue

Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

10/28

An open issue

► Popular syntax-driven approach to focus projection: Selkirk (1995) and many variants

► Recently, some authors have questioned whether there is an interesting constraining relation between syntax, intonation, and information structure:

   ► Büring (2006) any accent within a phrase can project focus to the phrase, i.e., focus can always project

   ► Kadmon (2006), Roberts (2006): Focus never projects!
     ► New elements must be accented.
     ► Unaccented focused elements are given/retrievable/expectable.

→ Makes fundamentally different empirical predictions than traditionally assumed.

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection

An open issue

Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

11/28

Experimental evidence

► There is very little published empirical research on focus projection.

► Perception experiments (Gussenhoven 1983; Birch & Clifton 1995; Welby 2003; Féry 1993).
   ► some evidence for the existence of focus projection in English, esp. for certain pairs of verbs with object-NPs

► Production experim. (Baumann et al. 2006; Féry & Kügler 2008)
   ► Significant variation in the use of a range of strategies (increased duration, height of peak, downstepping, . . .)

► More evidence is needed to determine:
   ► In which constructions can what kind of elements be accented (with which type of accents) and project focus how far?

► Can we obtain more robust empirical evidence for focus projection from authentic language data in corpora?

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection

An open issue

Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

12/28

The IMS Radionews Corpus

► Recordings of radio broadcasts on Deutschlandfunk: Total length: 1 h 26 min, 514 sentences (≈ 10 sec/sent)

► Corpus preparation involved (Rapp 1998):
   ► manual segmentation into news stories
   ► orthographic transliteration
   ► automatically word alignment
   ► manual prosodic labeling with ToBI

Slide 1 (top-left, 13/28)

Syntactic annotation of the corpus

- ▸ IMS Radionewscorpus is not syntactically annotated.
- ▸ Parsed corpus with Berkeley parser (Petrov & Klein 2007) to obtain syntactic analysis (thanks to Adriane Boyd).
- ▸ Converted corpus including
  - ▸ orthographic transcription
  - ▸ phonetic transcription
  - ▸ ToBI annotation
  - ▸ syntactic analysis

  into TiGer-XML so that it can be browsed and searched using TiGerSearch tool (Lezius 2002).

Focus Projection Between Theory and Evidence
Detmar Meurers
University of Tübingen

Motivation
Background
Expressing inf. struc
Discs. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence
IMS Radionews Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
An unexpected accent
Examples occurring twice
Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary
Conclusion

13/28

Slide 2 (top-right)

PP showing focus projection

6)

Bundesinnenminister Kanther hat sich
federal minister Kanther has self
L*H   L*H

gegen die Aufnahme weiterer Flüchtlinge aus Bosnien ausgesprochen
against the acceptance further refugees from Bosnia spoken
L*H   H*L   L%

Slide 3 (bottom-left, 15/28)

NP with multiple accents

7)

Der nordrhein-westfälische Ministerpräsident Rau hat den
the North Rhine Westphalian minister Rau has the
H*L   !H*L   L*H L%

Führungsstreit bei den Sozialdemokraten kritisiert.
leadership dispute among the social democrats criticized
H*L-   H*L   *? L%

Slide 4 (bottom-right, 16/28)

NP with multiple accents

Accents on every part of an NP 🔊

8)

Außenminister de Charette versicherte in dem heute von der Zeitung Sydney Morning Herald veröffentlichten Schreiben,

Foreign minister de Charette assured in a letter published by the newspaper Sydney Morning Herald today

von den Versuchen auf dem Mururoa-Atoll werde
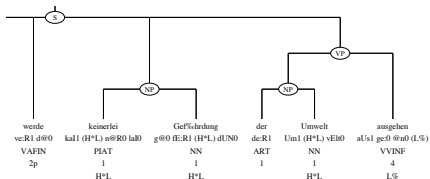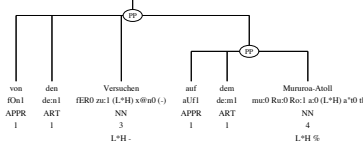of the testing   on the Mururoa atoll will
*L*H-   L*H%*

keinerlei Gefährdung der Umwelt   ausgehen.
no   harm   the environment emanate
*H*L   H*L   H*L   L%*

Accents on every part of an NP 🔊

An unexpected accented in an NP 🔊

9)

Der deutsche Sparkassen- und Giroverband hat davor gewarnt, die psychologischen und praktischen Probleme bei der Einführung einer gemeinsamen europäischen Währung zu unterschätzen. Die Konvergenzkriterien müßten unbedingt eingehalten werden, betonte Köhler in einem Interview. Bloße Tendenzen reichten dabei nicht aus,

The German banks warned that the psychological and practical problems with introducing the joint currency should not be underestimated. The convergence criteria must definitely be observed, said Köhler in an interview. Bare tendencies are not sufficient,

es dürfe nicht zu einer Verunsicherung der Menschen kommen.
it needs not to a   uncertainty   of the people   come
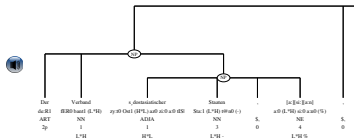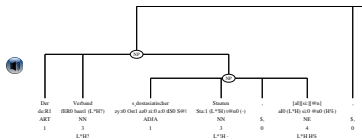*?   H*L   L%*

Examples occurring twice with different accents (1,2)

0)

Der Verband   südostasiatischer Staaten, ASEAN, hat heute auf
the organization southeast Asian nations ASEAN has today on
*L*H?   L*!H-   L*H H%   H*L?*
*L*H   H*L   L*H-   L*H%   H*L-*

seiner Jahrestagung im Sultanat Brunei Vietnam aufgenommen.
its   annual meeting in the sultanate Brunei Vietnam affiliated
L!H-   L*H-   H*L*
*L*H   L*H%   H*L*

Examples occurring twice with different accents (1,2)

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Discs. fct. of pitch accents
Relating struc. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

22 / 28

Verbmobil corpus

- spontaneous speech from dialog task in appointment scheduling domain
- subcorpus of 917 dialogues (1841 turns) was manually GToBI labeled
- parsed with Berkeley parser (Petrov & Klein 2007)
- Exploration of the Verbmobil corpus confirms patterns found for IMS Radionews corpus:
  - some examples illustrating focus projection patterns
  - others showing substantial additional accentuation
  - significant amount of variation in the realizations

Typical focus projection pattern

1)

Ja, Frau Petz, dann lassen Sie uns doch einen Termin ausmachen.
yes Mrs Petz then let you us still a date schedule
H*   H*   H*

- H* on noun *Termin* projects onto *then* sentence

Common broad focus pattern

2)

Wie sieht das bei Ihnen am Donnerstag aus?
What does your Thursday look like?

Da muß ich leider zu einem Treffen nach Köln.
there must I unfortunately to a meeting in Köln
H*   L+H*   L+!H*

- Unaccented verbal heads and unaccented adjuncts are commonly found in the corpus.

Focus Projection
Between Theory
and Evidence
Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Discs. fct. of pitch accents
Relating struc. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

24 / 28

Slide 1 (top-left)

Accents on all content words

Wenn Sie mir noch kurz erklären, wie ich zu Ihnen komme?
Could you briefly explain how I can find you?

Sie	*finden*	*mich*	*im*	*zweiten*	*Stock*	*in*	*Zimmer*	*zweihundert*	*drei*
you	find	me	on the	second	floor	in	room	two hundred	three
H*				L+H*	L+H*		H*	!H*	!H*

Slide 2 (top-right)

Broad focus with variation in downstep use

a.
Was kann ich für Sie tun?
What can I do for you?

b.	*In*	*unserem*	*Projekt*	*ist*	*unerwartet*	*ein*	*Problem*	*aufgetaucht.*
	in our		project	is	unexpectedly	a	problem	surfaced
		L+H*		H*			L+H*	

c.	*Wir*	*müssen*	*möglichst*	*schnell*	*eine*	*Besprechung*	*ansetzen.*
	we must		preferably	quick	a	meeting	arrange
		H*		!H*		!H*	

▸ Broad focus, with H* accents on most content words,
▸ but in line with production study (Baumann et al. 2006) variation in downstep (used only in second example)

Slide 3 (bottom-left)

Summary of corpus exploration

▸ 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.

Slide 4 (bottom-right)

Conclusion

▸ In light of conflicting theory-driven proposals on the interaction of syntax, intonation, and information structure, it is crucial to strengthen the empirical basis.
▸ Spoken corpora are attractive source of authentic data
  ▸ Targeted exploration requires annotated corpora.
  ▸ Where treebanks are not available, parsing is a viable option for supporting syntactic corpus queries.
▸ Reported on exploration based on parsed IMS Radionews Corpus and Verbmobil Corpus.
  ▸ Traditional focus projection patterns exist, yet only as part of an empirical landscape with rich variation.
▸ Larger intonationally annotated corpora of spoken language in context are needed to further explore the interaction of information structure, syntax, and intonation.

References

Baumann, S., M. Grice & S. Steindamm (2006). Prosodic Marking of Focus Domains - Categorical or Gradient? In *Proceedings SpeechProsody 2006*. Dresden, Germany, pp. 301–304.

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

Birch, S. & 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 & S. Winkler (eds.), *The Architecture of Focus*, Berlin: Mouton De Gruyter, vol. 82 of *Studies in Generative Grammar*, pp. 321–346.

Chomsky, N. & M. Halle (1968). *The Sound Pattern of English*. New York, NY: Harper and Row.

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. & 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*. No. 285 in Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag.

Féry, C. & F. Kügler (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4), 680–703.

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 Presentation in the Department of Linguistics Colloquium, OSU, October 19, 2006. (Based on a talk given at the Colloque de Syntax et Sémantique in Paris, 2000).

Lezius, W. (2002). Ein Suchwerkzeug für syntaktisch annotierte Textkorpora. Ph.D. thesis, IMS, University of Stuttgart. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), vol 8, nr 4.

Petrov, S. & D. Klein (2007). Improved Inference for Unlexicalized Parsing. In *HLT/ACL 2007*. pp. 404–411. URL http://www.aclweb.org/anthology/N/N07/N07-1051.

Rapp, S. (1998). Automatisierte Erstellung von Korpora für die Prosodieforschung. Ph.D. thesis, Univ. Stuttgart.

Roberts, C. (2006). Resolving Focus. Conference abstract for Sinn und Bedeutung 11. Barcelona, Spain. URL http://purl.org/dm/tmp/Roberts-Sinn-und-Bedeutung-2006.pdf.

Selkirk, E. (1995). Sentence Prosody: Intonation, Stress, and Phrasing. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford: Basil Blackwell, chap. 16, pp. 550–569.

Uhmann, S. (1991). *Fokusphonologie*. No. 252 in Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag.

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

Focus Projection
Between Theory
and Evidence

Detmar Meurers
University of Tübingen

Motivation

Background
Expressing inf. struc
Disco. fct. of pitch accents
Relating inton. & interpretation
Focus projection
Constraining projection
An open issue
Experimental evidence

IMS Radionews
Corpus
Syntactic corpus annotation
PP showing focus projection
NP with multiple accents
Accents on every part
An unexpected accent
Examples occurring twice

Verbmobil Corpus
Focus projection example
Common broad focus pattern
Accents on all contents words
Variation in downstep use
Corpus exploration summary

Conclusion

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

28 / 28