Studying the prosodic properties of referring expressions in corpus resources: obstacles & state of the art

AG2: The relation between prosodic & referential structure

Universität Stuttgart

A. Riester, K. Eckart, M. Gärtner, I. Kasimir, I. Rösiger, K. Schweitzer, A. Schweitzer, S. Stehwien (SFB 732)

DGfS – March 7, 2018
Overview

1. Introduction
   - Pragmatics and prosody
   - Previous study on read speech

2. DIRNDL corpus: German radio news

3. German interviews: GRAIN
   - PaIntE-based classification of GToBI(S) events
   - Information status revisited
   - QUDs and information structure
   - Results
Introduction
Pragmatics and prosody

• The prosodic realization of spoken utterances depends on discourse context.

• English, German (+ many other languages): Alignment between focus constituent and nucleus of intonation phrase [Truckenbrodt, 1995, Büring, 2016]

• Focus:
  • Answer to a contextual question (QUAD) [Roberts, 2012]
  • thus, typically new information
  • but also contrastive / choice-related / re-activating etc. (still very much dependent on terminology & theory!)

• Information status of referring expressions (terms, markables):
  • Given terms have a stronger tendency to become backgrounded (topical) than new terms.
  • New terms have a stronger tendency to become focal than given terms.
Information status vs. focus annotation

- Information status [Prince, 1981, Prince, 1992] of referring expressions is a simpler concept than focus-background structure of full utterances
- Info status is easier to annotate than focus:

<table>
<thead>
<tr>
<th>Text type</th>
<th>Info status given/acc/new (NP-based)</th>
<th>Info status 7 categories (NP-based)</th>
<th>Focus (token-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question/answer</td>
<td>.80</td>
<td>.73</td>
<td>.51</td>
</tr>
<tr>
<td>Dialogue</td>
<td>.66</td>
<td>.61</td>
<td>.44</td>
</tr>
<tr>
<td>News commentary</td>
<td>.60</td>
<td>.55</td>
<td>.19</td>
</tr>
</tbody>
</table>

κ-values [Dipper et al., 2007, Ritz et al., 2008]
Pitch accent types and German referring expressions

- Heterogeneous picture
  

  - Some reported tendencies that newness is marked with H* or Early Peak (H+L*, H+!H*) accents,
  - accessibility with downstepped !H* or Early Peak
  - givenness with L* (unless deaccented)
  - Studies show that all information status classes occur in all positions (prenuclear, nuclear, postnuclear) and with all accent types.

- Different GToBI schemes, [Mayer, 1995, Grice et al., 2005, Kügler et al., 2015] etc., combined with different information status classifications, are likely to yield widely diverging results.
Prosody of German referring expressions

[Baumann and Riester, 2013]

- Domains: read speech (controlled episodes) and spontaneous monologue
- Impact of two different levels of information status on the prosodic realization of short referring expressions
  - Accent position within IP: prenuclear, nuclear, phrase accent, deaccentuation
  - Accent type: H*, !H*, Early Peak (H+L*), L*, none
**RefLex scheme** [Baumann and Riester, 2012]

Simplified:

### Referential level (NPs)

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-GIVEN</td>
<td>coreference anaphor</td>
</tr>
<tr>
<td>R-BRIDGING</td>
<td>new but context-dependent</td>
</tr>
<tr>
<td>R-UNUSED</td>
<td>new definite</td>
</tr>
<tr>
<td>R-GENERIC</td>
<td>generic definite / indefinite</td>
</tr>
<tr>
<td>R-NEW</td>
<td>specific indefinite</td>
</tr>
<tr>
<td>OTHER</td>
<td>e.g. cataphors</td>
</tr>
</tbody>
</table>

### Lexical level (nouns, verbs, adjectives, adverbs)

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-GIVEN</td>
<td>repetition, synonym, hypernym, holonym</td>
</tr>
<tr>
<td>L-ACCESSIBLE</td>
<td>hyponym, meronym</td>
</tr>
<tr>
<td>L-NEW</td>
<td>no relation within previous 5 clauses</td>
</tr>
</tbody>
</table>

Inter-annotator agreement on radio news: [Riester and Baumann, 2013]

- R-LEVEL (6 categories) \( \kappa = .75 \)
- L-LEVEL (3 categories) \( \kappa = .64 \)
Results: read speech

Material:

- Ten stories labelled according to core RefLex scheme
- Each text read by 10 speakers (7f, 3m) in a natural manner
- Total of 450 sentences per speaker
- Prosodic labelling according to GToBI [Grice et al., 2005]
- Results reported are based on 807 referential expressions
### Accent positions (within prosodic phrase) in read speech

<table>
<thead>
<tr>
<th>Reflex combinations</th>
<th>nuclear accent</th>
<th>prenuclear accent</th>
<th>phrase accent</th>
<th>no accent</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>r-new / l-new</td>
<td>68</td>
<td>27</td>
<td>3</td>
<td>3</td>
<td>102</td>
</tr>
<tr>
<td>r-unused / l-new</td>
<td>24</td>
<td>71</td>
<td>3</td>
<td>2</td>
<td>260</td>
</tr>
<tr>
<td>r-generic / l-new</td>
<td>34</td>
<td>56</td>
<td>1</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td>r-bridging / l-new</td>
<td>28</td>
<td>47</td>
<td>2</td>
<td>23</td>
<td>57</td>
</tr>
<tr>
<td>r-given / l-new</td>
<td>37</td>
<td>33</td>
<td>9</td>
<td>21</td>
<td>67</td>
</tr>
<tr>
<td>r-bridging / l-accessible</td>
<td>60</td>
<td>17</td>
<td>9</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td>r-new / l-given</td>
<td>11</td>
<td>78</td>
<td>11</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>r-bridging / l-given</td>
<td>29</td>
<td>25</td>
<td>33</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>r-given / l-given</td>
<td>9</td>
<td>23</td>
<td>16</td>
<td>51</td>
<td>128</td>
</tr>
</tbody>
</table>

**Distribution of accent positions (%)**
## Accent types in read speech

![Bar chart showing distribution of accent types](chart.png)

<table>
<thead>
<tr>
<th>Reflex combinations</th>
<th>H*</th>
<th>Ĥ*</th>
<th>EP</th>
<th>L*</th>
<th>No pitch accent</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>r-new / l-new</code></td>
<td>19</td>
<td>19</td>
<td>35</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td><code>r-unused / l-new</code></td>
<td>43</td>
<td>8</td>
<td>5</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td><code>r-generic / l-new</code></td>
<td>30</td>
<td>12</td>
<td>14</td>
<td>34</td>
<td>10</td>
</tr>
<tr>
<td><code>r-bridging / l-new</code></td>
<td>28</td>
<td>4</td>
<td>9</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td><code>r-given / l-new</code></td>
<td>22</td>
<td>3</td>
<td>13</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td><code>r-bridging / l-accessible</code></td>
<td>9</td>
<td>3</td>
<td>47</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td><code>r-new / l-given</code></td>
<td>6</td>
<td>17</td>
<td>67</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td><code>r-bridging / l-given</code></td>
<td>4</td>
<td>38</td>
<td>13</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td><code>r-given / l-given</code></td>
<td>16</td>
<td>4</td>
<td>13</td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>102</th>
<th>280</th>
<th>93</th>
<th>57</th>
<th>67</th>
<th>58</th>
<th>18</th>
<th>24</th>
<th>128</th>
</tr>
</thead>
</table>

*Note: n represents the number of occurrences.*
Results: read speech (cont.)

- Weighting of accent types (H* > !H* > EP > L*> none) and accent positions (nuclear > prenuclear > phrase accent > deaccented)
- Both levels of information status have an effect on prosodic prominence

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Accent pos.</th>
<th>Accent type</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-NEW/L-NEW &gt; R-GIVEN/L-GIVEN</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>R-NEW/L-NEW &gt; R-NEW/L-GIVEN</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>R-NEW/L-NEW &gt; R-GIVEN/L-NEW</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>R-GIVEN/L-NEW &gt; R-GIVEN/L-GIVEN</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>R-NEW/L-GIVEN &gt; R-GIVEN/L-GIVEN</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

- Results could not be confirmed on spontaneous monologues.
DIRNDL corpus: German radio news
[Eckart et al., 2012, Björkelund et al., 2014]  
http://www.ims.uni-stuttgart.de/data/dirndl/

- Spoken radio news (*Deutschlandfunk 25-27/03/2007*)
- 5 hrs., 50,000 tokens, 3,221 sentences, 9 speakers (4f, 5m)

### Annotation layers

**Manual:**
- Prosody: GToBI(S) [Mayer, 1995]
- Pragmatics: RefLex [Baumann & Riester, 2012]

**Automatic:**
- Prosody: PaIntE [Möhler, 2001]
- Morphology:
  - Lemmata [Bohnet, 2010]
- POS & morphology [Mueller et al., 2013]

**Syntax:**
- constituent trees [Rohrer & Forst 2006]
- constituent & dependency trees [Björkelund et al., 2013]

**Semantics:**
- named entities [Finkel et al., 2005]

- PostgreSQL database
- DIRNDL*anaphora* available in CoNLL shared task format
- Used to train coreference resolver with prosodic features

[Rösiger and Riester, 2015, Rösiger et al., 2017]
DIRNDL: linking annotation graphs
GToBI(S)

[Mayer, 1995]

- German ToBI, Stuttgart “dialect”
- boundary tones and breaks: %, H%, L%, -
- possibly downstepped
Prosody of referring expressions in DIRNDL

[Kasimir, in prep.]

- Extracted 9056 reference phrases (R-phrases)
- Pronouns excluded
- Under scrutiny: prosody of phrase-final tokens
- Different labelling scheme: GToBI(S) [Mayer, 1995]

<table>
<thead>
<tr>
<th>Main class</th>
<th>Labels in DIRNDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>H*</td>
<td>H*, H*+L</td>
</tr>
<tr>
<td>!H*</td>
<td>!H*, !H*+L</td>
</tr>
<tr>
<td>L*</td>
<td>L*, L*+H</td>
</tr>
</tbody>
</table>
Accent types in DIRNDL

![Bar chart showing distribution of accent types](image-url)
German interviews: GRAIN
GRAIN (German Radio Interviews)

[Eckart and Gärtner, 2016, Schweitzer et al., 2018]
http://hdl.handle.net/11022/1007-0000-0007-C632-1

- German radio interviews (SWR2 Interview der Woche)
- Part of SFB732 Silver Standard Collection

The silver-standard idea [Rebholz-Schuhmann et al., 2010]

- Verified automatic annotation by combining annotations for the same layer
- Confidence estimations
- Bigger resources than by (manual) gold annotations

- Use of state-of-the-art tools for text and speech processing
- Spontaneous speech by experienced public speakers
- Challenge for text-processing tools
GRAIN: Primary data

- Each interview just under 10 min, mp3
- Edited transcripts from radio station
- 3 interviews chosen for training annotators and automatic tools
- 20 interviews chosen for gold annotations
- Non-gold part growing as radio station releases more interviews (currently 144 interviews, about 221,000 word tokens, 23 hrs audio)
- Thorough workflow documentation
GRAIN: Automatic annotation

- Pre-processing: speaker turns, document structure
- Tokenization [Schmid, 1994], sentence segmentation
- Acoustic alignment [Rapp, 1998]: phone, word, syllable boundaries
- Parametrized Intonation Events (PaIntE) [Möhler, 2001]: parameters describing $F_0$ shape
- Intonation:
  - PaIntE-based prediction of intonation events [Schweitzer, 2010]: GToBI(S) pitch accents and boundary tone types
  - CNN-based prediction of pitch accent placement [Stehwien and Vu, 2017]
- Additional phonetic features for each syllable
- Morpho-syntax:
  - 3 constituency, 4 dependency parsers
  - with underlying morpho-syntactic annotations
  - Confidence estimations as meta-annotation (based on agreement of different tools)
GRAIN: Manual annotation

- Re-introducing some features of orality [Eckart and Gärtner, 2016]
- (Additional) part-of-speech tagging [Seeker, 2016]
- Referential information status [Riester and Baumann, 2017]
- Questions under Discussion, discourse structure and information structure [Reyle and Riester, 2016, Riester et al., to appear]
PaIntE-based classification of GToBI(S) events

Methodology [Schweitzer, 2010]

- SmartWeb speech synthesis database
- Professional male speaker
- Approx. 2 hrs., 28,000 syllables
- Read speech, mostly news-like style
- Automatic aligned at segment, syllable, word levels
- Manually GToBI(S) annotated [Mayer, 1995]
Attributes for classifying prosodic categories

- 20 parameters related to PaIntE (including context syllables)
- 6 parameters related to duration z-scores (including context syllables)
- 10 higher-linguistic attributes including word stress, silences, part-of-speech tags, etc.
## Word-based accuracies for best algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>accents</th>
<th>boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-class</td>
<td>full set</td>
</tr>
<tr>
<td>Bagging</td>
<td>86.19</td>
<td>78.08</td>
</tr>
<tr>
<td>ClassificationViaRegression</td>
<td>85.49</td>
<td>78.17</td>
</tr>
<tr>
<td>LogisticModelTree</td>
<td>86.24</td>
<td>77.54</td>
</tr>
<tr>
<td>RandomForest</td>
<td>86.17</td>
<td>78.04</td>
</tr>
<tr>
<td>ZeroR $\hat{=}$ Baseline</td>
<td>58.30</td>
<td>58.23</td>
</tr>
</tbody>
</table>

Details see [Schweitzer, 2011](#)
Illustration of results

manual

auto

IMS Stuttgart, Prosody, Reference, Corpora
Illustration of results

<table>
<thead>
<tr>
<th>word</th>
<th>accents</th>
<th>boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>correct</td>
<td>predicted</td>
</tr>
<tr>
<td>Das Zentrum</td>
<td>L*HL</td>
<td>L*HL</td>
</tr>
<tr>
<td>blieb</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>zumeist</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>Cardoso</td>
<td>H*L</td>
<td>H*L</td>
</tr>
<tr>
<td>vorbehalten</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>weil</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>Spörl</td>
<td>L*H</td>
<td>L*H</td>
</tr>
<tr>
<td>sich</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>deutlich</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>rechts</td>
<td>H*L</td>
<td>NONE</td>
</tr>
<tr>
<td>orientierte</td>
<td>NONE</td>
<td>NONE</td>
</tr>
</tbody>
</table>
Refinement of RefLex guidelines

[Riester and Baumann, 2017]

### Referential level (NPs)

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-GIVEN-SIT</td>
<td>deictic term</td>
</tr>
<tr>
<td>R-GIVEN</td>
<td>coreference anaphor</td>
</tr>
<tr>
<td>R-GIVEN-DISPLACED</td>
<td>antecedent &gt; 5 clauses</td>
</tr>
<tr>
<td>R-CATAPHOR</td>
<td>cataphor</td>
</tr>
<tr>
<td>R-BRIDGING</td>
<td>new but context-dependent</td>
</tr>
<tr>
<td>R-BRIDGING-CONTAINED</td>
<td>bridging anaphor containing antecedent</td>
</tr>
<tr>
<td>R-UNUSED-KNOWN</td>
<td>new familiar definite</td>
</tr>
<tr>
<td>R-UNUSED-UNKNOWN</td>
<td>new unfamiliar definite</td>
</tr>
<tr>
<td>R-NEW</td>
<td>specific indefinite</td>
</tr>
<tr>
<td>R-EXPLETIVE</td>
<td>pleonastic pronoun</td>
</tr>
<tr>
<td>R-IDIOM</td>
<td>idiomatic term</td>
</tr>
</tbody>
</table>
Inter-annotator agreement (11 categories on GRAIN)

<table>
<thead>
<tr>
<th>Annotators</th>
<th>#Documents</th>
<th>#Markables</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+B</td>
<td>5</td>
<td>1808</td>
<td>0.822</td>
</tr>
<tr>
<td>A+C</td>
<td>5</td>
<td>1917</td>
<td>0.788</td>
</tr>
<tr>
<td>A+D</td>
<td>5</td>
<td>1608</td>
<td>0.734</td>
</tr>
<tr>
<td>A+E</td>
<td>4</td>
<td>1181</td>
<td>0.654</td>
</tr>
<tr>
<td>B+C</td>
<td>6</td>
<td>2282</td>
<td>0.759</td>
</tr>
<tr>
<td>B+D</td>
<td>3</td>
<td>775</td>
<td>0.696</td>
</tr>
<tr>
<td>B+E</td>
<td>3</td>
<td>1042</td>
<td>0.635</td>
</tr>
<tr>
<td>C+D</td>
<td>2</td>
<td>510</td>
<td>0.651</td>
</tr>
<tr>
<td>C+E</td>
<td>1</td>
<td>243</td>
<td>0.712</td>
</tr>
<tr>
<td>D+E</td>
<td>6</td>
<td>2038</td>
<td>0.733</td>
</tr>
<tr>
<td>A+B+C</td>
<td>3</td>
<td>825</td>
<td>0.783</td>
</tr>
<tr>
<td>A+B+D</td>
<td>2</td>
<td>383</td>
<td>0.711</td>
</tr>
<tr>
<td>A+B+E</td>
<td>1</td>
<td>173</td>
<td>0.756</td>
</tr>
<tr>
<td>A+C+D</td>
<td>2</td>
<td>417</td>
<td>0.707</td>
</tr>
<tr>
<td>A+C+E</td>
<td>1</td>
<td>196</td>
<td>0.753</td>
</tr>
<tr>
<td>A+D+E</td>
<td>2</td>
<td>235</td>
<td>0.775</td>
</tr>
<tr>
<td>B+C+D</td>
<td>2</td>
<td>390</td>
<td>0.688</td>
</tr>
<tr>
<td>B+C+E</td>
<td>1</td>
<td>176</td>
<td>0.737</td>
</tr>
<tr>
<td>A+B+C+D</td>
<td>2</td>
<td>352</td>
<td>0.729</td>
</tr>
<tr>
<td>A+B+C+E</td>
<td>1</td>
<td>153</td>
<td>0.777</td>
</tr>
<tr>
<td>B+C+D+E</td>
<td>1</td>
<td>171</td>
<td>0.742</td>
</tr>
<tr>
<td>A+B+C+D+E</td>
<td>1</td>
<td>149</td>
<td>0.770</td>
</tr>
</tbody>
</table>

- [Pagel, 2018]
- Kappa statistics for all annotator pairs over all categories
- Number of shared documents, number of markables with same span and Fleiss’ Kappa values for all annotator pairs over all categories.

$\kappa$ average: **0.75**
QUD trees

- Method for the simultaneous analysis of discourse structure & information structure (focus-background)
- Assertions: terminal nodes (in linear order)
- Questions: non-terminal nodes

[Reyle and Riester 2016, Riester et al., to appear]
What is hidden behind the text

Every assertion in a text is the answer to an (implicit) question.

Focus (F) Answer to the QUD
Background (BG) Material contained in QUD
Focus domain (~) Phrase that contains a focus; establishes question-answer congruence

Contrastive topic (CT) Focus-like indicator of a subquestion [Büring, 2003]

Non-at-issue mat. (NAI) Optional material, which does not answer the QUD

(1) \[ Q_1: \{ \text{Who came home when?} \} \]
   \[ > Q_{1.1}: \{ \text{When did George come home?} \} \]
   \[ >> A_{1.1}: [\text{I believe that,}]_\text{NAI} [[\text{George}]_\text{CT} \text{ came home [at eight]}]_\sim. \]
Quick guide: QUD-based discourse analysis

(Re)constructing QUDs – Givenness

[Schwarzschild, 1999]

“GIVENNESS: If a constituent is not F-marked, it must be GIVEN.”

If F-marked constituents are answers, then non-F-marked constituents must be part of the question background.

⇒ Q-GIVENNESS: (Implicit) QUDs can only consist of given material.

[Reyle and Riester, 2016]
Quick guide: QUD-based discourse analysis

**Q-Givenness:** QUDs can only consist of given material


(2) $Q_{3a}$: # \{Who has been bankrupt already since 2010?\}  
$Q_{3b}$: # \{What has Greece been already since 2010?\}  
$Q_{3c}$: \{Since when has Greece been bankrupt?\}

Other principles

**Maximize-Q-Anaphoricity**
Implicit QUDs must contain as much given (or salient) material as possible.

**Parallelism**
A QUD with two or more parallel answers contains the (semantically) shared material of the answers.

Parallelism can override Q-Givenness.
Sample annotation

A_{23}: Und man muss jetzt aufpassen, dass man sich nicht zum Sprachrohr von Leuten macht, die eben den Mindestlohn umgehen wollen.

Q_{27}: {Was sind die Hauptstreitpunkte?}

A_{27}: [[Einer der Hauptstreitpunkte]T ist ja [die Dokumentationspflicht.]F]~

Q_{28}: {Was bemängelt die CDU bei der Dokumentationspflicht?}

Q_{35}: {Was ist Nahles' Reaktion auf den Streit?}

Q_{36}: {Wie kann man die Dokumentation gestalten?}


A_{36}#: [[das]T kann [man]T machen [in jedweder Form.][F]~

A_{26}: [[Die Union.]T die zielt [da]T immer wieder auf [diese Einkommensgrenze von 2958 Euro.]F]~

Q_{29}: {Was ist mit der Grenze?}

A_{29}: [bis zu [der]T [in bestimmten Branchen]NAl [die Arbeitszeiten genau erfasst werden müssen.]F]~

A_{36}: [Das T kann [man]T machen [in jedweder Form.]F]~

SWR2 Interview der Woche (28.02.2015) with Andrea Nahles
TreeAnno: QUD-annotation, visualization, evaluation, export

[De Kuthy et al., 2018]
Evaluation

Discourse structure:
- Trees transformed into matrix containing Q/none
- Cohen’s $\kappa$ based on [Marcu et al., 1999]

<table>
<thead>
<tr>
<th>Text type</th>
<th>Interview (GER)</th>
<th>Interview (ENG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ (cells)</td>
<td>.45</td>
<td>.50-.53</td>
</tr>
</tbody>
</table>

Information structure:

<table>
<thead>
<tr>
<th>Guidelines Evaluation Evaluation Categories Text Type $\kappa$ (token)</th>
<th>[Dipper et al., 2007]</th>
<th>[Ritz et al., 2008]</th>
<th>Riester et al. (to appear)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F, NONE</td>
<td>Que/Ans</td>
<td>Interv. (GER) Interv. (ENG)</td>
</tr>
<tr>
<td></td>
<td>.51</td>
<td>.44</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.19</td>
<td>.67-.69</td>
</tr>
</tbody>
</table>
Pitch accents and referring expressions in GRAIN

- First results based on 4 annotated interviews
GRAIN: information status and nuclear pitch accents on short
referring phrases (<= 3 tokens)
ICARUS

[Gärtner et al., 2015, Schweitzer et al., 2015]

http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/icarus.html

- Search and visualization tool for text and speech
- Freely available, open source Java application
- Query corpora with textual and prosodic annotations
GRAIN: information status vs. all pitch accents on short referring phrases (≤ 3 tokens)
GRAIN: information structure vs. pitch accent types on short phrases
GRAIN: Focus/Background, Given/New and pitch accent type

Nuclear H*L and L*H accents

- **H*L
- **L*H

Frequency

**Background**

**Focus**

**R−GIVEN**

**R−NEW**

IMS Stuttgart, Prosody, Reference, Corpora
Outlook

• More data available soon
  • 4 doc. → 13 doc. (Focus & InfoStatus)
  • → 23 doc. (InfoStatus)
• Second labeler for pitch accents involving CNNs
• Assessment of automatic intonation and phonetic parameters
• Syntactic complexity of the phrases
• Location for default accent rather than last token
THANK YOU!

Credits

Kerstin Eckart  corpus infrastructure
Katrin Schweitzer  speech infrastructure, statistics
Antje Schweitzer  speech pipeline, automatic prediction of intonation
Markus Gärtner  corpus infrastructure, ICARUS
Ina Rösiger  information status management
Ingrid Kasimir  DIRNDL statistics
Sabrina Stehwien  accent prediction with neural networks
Janis Pagel  information status annotation, inter-annotator agreement
Daniel Knaus  information structure annotation
Nadja Schäfer  information structure annotation
Anne-Cathrine Draudt  inter-annotator agreement

This research was supported by the SFB 732 Incremental Specification in Context (University of Stuttgart), funded by the German Research Foundation (DFG).
References I

  The intonation of accessibility.

  Referential and lexical givenness: Semantic, prosodic and cognitive aspects.

  Coreference, lexical givenness and prosody in german.

  Questions Under Discussion: Where information structure meets projective content.

  (Re)ranking meets morphosyntax: State-of-the-art results from the SPMRL 2013 shared task.

  The extended DIRNDL corpus as a resource for automatic coreference and bridging resolution.

  Top accuracy and fast dependency parsing is not a contradiction.

  On D-trees, beans, and B-accents.
References II

Intonation and Meaning.
Oxford University Press.

QUD-based annotation of discourse structure and information structure: Tool and evaluation.
In Proceedings of the 11th Language Resources and Evaluation Conference (LREC), Miyazaki, JP.

Information status marking in spontaneous vs. read speech in story-telling tasks – Evidence from intonation analysis using (gtobi).

Universitätsverlag Potsdam.

Creating silver standard annotations for a corpus of non-standard data.
In Dipper, S., Neubarth, F., and Zinsmeister, H., editors, Proceedings of the 13th Conference on Natural Language Processing (KONVENS), pages 90–96, Bochum, Germany.

A Discourse Information Radio News Database for Linguistic Analysis.

Incorporating non-local information into information extraction systems by Gibbs sampling.
In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, pages 363–370, Ann Arbor, MI, USA.
References III

Multi-modal visualization and search for text and prosody annotations.

German intonation in Autosegmental-Metrical Phonology.

The Phonology of Tone and Intonation.
Cambridge University Press.

Analysing the effect of information status on prosody in a German radio news corpus.
in preparation.

Quaestio und referentielle Bewegung in Erzählungen.

Form and function of non-pitch accents.
Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK), 35a:97–123.

DIMA: Annotation guidelines for German intonation.
In Wolters, M. et al., editors, Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), Glasgow.
References IV


References V

Automatisierte Erstellung von Korpora für die Prosodieforschung.

The calbc silver standard corpus for biomedical named entities: A study in harmonizing the contributions from four independent named
entity taggers.
In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta.

Joint information structure and discourse structure analysis in an Underspecified DRT framework.
(JerSem), pages 15–24, New Brunswick, NJ, USA.

Focus triggers and focus types from a corpus perspective.
Dialogue and Discourse, 4(2).

The RefLex Scheme – Annotation Guidelines, volume 14 of SinSpeC. Working Papers of the SFB 732.
University of Stuttgart.

Annotation guidelines for Questions under Discussion and information structure.
In Evangelia Adamou, K. H. and Vanhove, M., editors, Information Structure in Lesser-Described Languages: Studies in Prosody and

Annotation of information structure: An evaluation across different types of texts.
In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC), pages 2137–2142, Marrakesh.
References VI

Information structure: Towards an integrated formal theory of pragmatics.

Decoding information status by type and position of accent in German.
*Proceedings 17th International Congress of the Phonetic Sciences (ICPhS)*, pages 1706–1709.

Improving Coverage and Parsing Quality of a Large-Scale LFG for German.
In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, Genova.

A Theory of Focus Interpretation.

Using prosodic annotations to improve coreference resolution of spoken text.

Improving coreference resolution with automatically predicted prosodic information.
In *Proceedings of the First Workshop on Speech-Centric Natural Language Processing*, pages 78–83, Copenhagen.

Probabilistic part-of-speech tagging using decision trees.

GIVENness, AvoidF, and other constraints on the placement of accent.
*Natural Language Semantics*, 7(2):141–177.
References VII


