Annotating Information Structure in the Old High German Tatian translation

Svetlana Petrova
Humboldt University Berlin

Corpus Meeting Berlin
November 13, 2009
Subjects of investigation in Project B4, SFB 632

- Is IS responsible for syntactic variation in early Germanic?
  Is there a correlation between the pragmatic properties of constituents and their position in the clause?

- How does IS influence language change?
  Which factors influence the emergence of novel word order patterns and the generalization of originally marked word orders?

- Empirical investigations on the following phenomena:
  - The (non-)development of generalized V2 in root declarative clauses
  - The variation between OV and VO in Germanic

Focus on Old High German
A brief note on the quality of the data

- no authentic OHG prose texts, only translations and poetry
- the attested word order patterns and constructions may not be representative for native OHG grammar but influenced by the syntax of the original or by metrical considerations

  - OHG translations provide native evidence in cases where the vernacular text changes the constituent order of the corresponding Latin clause
  - the largest corpus of potential native structures can be obtained from the OHG Tatian (340 fol.), the bilingually attested Gospel Harmony of Tatian (Lat.-OHG)
Exsurgen Autem Ioseph Annono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.

Le.

Exsurgen Autem Ioseph Antiono
famnet nuptiis priepere angeludni
s'accepto coiusugem suam
s'amoseognorabae tam donee pepere
filium suum primum genenrum.

S'acceper nomen eur ibi.

Faciam eur Autem Indeburillo.
Focus on non-Latin orders as potential native patterns („Differenzbelege“)

- discourse referent=new
- inflected verb
Focus on non-Latin orders as potential native patterns („Differenzbelege“)

- insertion of constituents (su/obj pronouns, adverbs, copulae)
  (1) *eo quod ess& de domo / & familia dauid*  
  
  bithiu uuanta *her uuas fon huse / Inti fon hiuuiske dauides*
  
  because he was of the house and the lineage of D. (35,18f)

- reordering of constituents
  (2) *& ait angelus ei*  
  
  *quad Iru ther engil*
  
  spoke to her the engel (28,15)

- freely built clauses without a Latin counterpart
  (3) *de ciuitate nazar&h*  
  
  *fon theru burgi thiu hiez nazar&h*
  
  from the city that was called N. (35,15)
Levels of IS-relevant features

- informational status
  - theme/given vs. rheme/new

- predicational separation
  - topic vs. comment

- informational relevance
  - focus vs. background

(Molnár 1993, Krifka 2007)
Informational status

given = an explicitly mentioned antecedent is present in the context
(new) = no active antecedent at this point in the discourse

➢ no longer regarded as a dichotomy but rather as a continuum

accessibility = referents are inferable without explicit mention

❖ situational accessibility: items of the extra-linguistic context
❖ anchoring: a friend of mine, a woman I work with
❖ bridging (Sub-Topic): bus-driver; party-food/music
❖ world knowledge: proper names, geographic or theological concepts

Informational status

- restrictions: givenness is only properly applicable to discourse referents (DR)

- Karttunen (1976): DR are individuals which can be taken up by coreferent pronouns in the subsequent context

(4) I have a new car. It is blue.

(5) I have no car. *It is blue.

- definite expressions denoting individuals (proper names, DPs, referring pronouns)
  - but not
- quantified and negative expressions (each X, every X, no X...)
- nominal predicates (*Peter is an excellent researcher*)
- nominal parts of idioms (*My uncle kicked the bucket*)
Informational status

➤ exceptions:

❖ short-time referents in modal and negative contexts

(6) I want to catch a fish.
   a. *It is big.
   b. I want to cook it.

❖ nominal parts of idioms

(7) Besser ein Spatz in der Hand als ein Taube auf dem Dach,
   'a bird in the hand is worth two in the bush`

→ in a special discourse situation the nominal part clearly refers to a
   previously introduced entity in the discourse (Chiarcos, p.c.)
Topic vs. Comment: some famous definitions

- topics as frame-setters
  
  "topics set a spatial, temporal, or individual framework within which the main predication holds" (Chafe 1976, 50)

- the aboutness concept
  
  "the speaker announces a topic and then says something about it"    Hockett (1958, 201)

- the familiarity concept
  
  "An entity E can ... serve as a topic ... iff, both speaker and addressee have previous knowledge of or familiarity with E"    (Gundel 1988, 212)
Restrictions for topicality

- referentiality
  
  "NP sentence-topics [are] referential entries under which we classify propositions"  
  (Reinhart 1981, 80)

  → only NPs with referential properties, but not negative or quantified expressions can act as suitable representatives for X in topic-tests of the type

  (8) As for X, X... or (9) A says about X that X...

- topic-comment and the thetic/categorical distinction

  → only categorical utterances involve a "bipartite division into a predication base, or topic, and a comment on this topic"

  Sasse (1995, 4)

  (10) A: What’s the matter?  (11) A: How’s your neck?

  B: My NECK hurts.  B: [My neck]_{TOP} HURTS.
Cumulative Approach: topic-related features

- Topic-related features:
  - individual denoting expressions (proper names, referential pronouns and full DPs)
  - given / identifiable entities in the context
  - definite DPs
  - subject of the predication in categorical utterances (topic test is applicable)
  - Topic marking constructions, e.g. German: left dislocation, hanging topic
Informational Relevance: Focus vs. Background

- Focus reflects the most relevant information in an utterance
  - this information is explicitly requested in a question: completive, or new-information focus
  - this information is in a relation of contrast to another part of the utterances: contrastive focus
- Both types of focus are compatible with the idea that focus expresses the presence of alternatives (Roots 1992)
- problem: new information focus in running narratives
- suggestion: try to determine the set of alternatives
- quaestio-theory: each sentence in a coherent discourse provides the answer to an implicite question which is unique at this point in the discourse (Klein & von Stutterheim 1992)
Implementation

- Design of an annotation scheme
- Annotation tool: Exmaralda allowing for multi-level annotation
- Search program: ANNIS

A Linguistic Database for Annotated Information Structure (SFB 632, Project D1)
# The Annotation Scheme

<table>
<thead>
<tr>
<th>Latin</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OHG</td>
<td></td>
</tr>
<tr>
<td>informational status</td>
<td>given=giv; new=new; accessible=acc</td>
</tr>
<tr>
<td>TCS</td>
<td>top=topic; comm=comment</td>
</tr>
<tr>
<td>aboutness</td>
<td>aboutness=ab: X fits to topic-tests</td>
</tr>
<tr>
<td>definiteness</td>
<td>def=definite; indef=indefinite</td>
</tr>
</tbody>
</table>
| position       | init vs. noninit (in Tatian 1.0)  
                             X is involved in a special type of construction (Tatian 2.0) |
| FBS            | foc=focus; bgr=background |
| nif            | new-information focus |
| cf             | contrastive focus |
### Example 1

Lc 1,5: Fuit in diebus Herodis regis ... quidam sacerdos ... & uxor illi

OHG  siu uuarun rehtiu beidu fora gote 'they were righteous both for God'

<table>
<thead>
<tr>
<th>Latin</th>
<th>erant</th>
<th>autem</th>
<th>iusti</th>
<th>ambo</th>
<th>ante</th>
<th>deum</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHG</td>
<td>siu</td>
<td>uuarun</td>
<td>rehtiu</td>
<td>beidu</td>
<td>fora</td>
<td>gote</td>
</tr>
<tr>
<td>inf status</td>
<td>giv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>acc</td>
</tr>
<tr>
<td>TCS</td>
<td>top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aboutness</td>
<td>ab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>definiteness</td>
<td>def</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def</td>
</tr>
<tr>
<td>position</td>
<td>init</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>noninit</td>
</tr>
<tr>
<td>FBS</td>
<td>bgr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foc</td>
</tr>
<tr>
<td>nif</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>nif</td>
</tr>
<tr>
<td>cf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 2

Lc 2, 8: Et pastores eratn In regione eadem uigilantes & custodientes uigilias noctis

OHG  uuarun thô hirta In thero lantskeffi ‚there were shepherds in that region‘

<table>
<thead>
<tr>
<th>Latin</th>
<th>Et</th>
<th>pastores</th>
<th>erant</th>
<th>In</th>
<th>regione</th>
<th>eadem</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHG</td>
<td>uuarun</td>
<td>thô</td>
<td>hirta</td>
<td>In</td>
<td>thero</td>
<td>lantskeffi</td>
</tr>
<tr>
<td>inform status</td>
<td>new</td>
<td>giv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aboutness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>definiteness</td>
<td></td>
<td>indef</td>
<td>def</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>position</td>
<td></td>
<td>noninit</td>
<td></td>
<td>noninit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBS</td>
<td></td>
<td></td>
<td>foc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nif</td>
<td></td>
<td></td>
<td>nif</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Example 3

Lc 1: In menso autem sexto missus est angelus gabriel … ad uirginem...

OHG quad Iru ther engil (T 28, 15)

<table>
<thead>
<tr>
<th>Lat. &amp; ait</th>
<th>angelus</th>
<th>ei</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHG</td>
<td>quad</td>
<td>Iru ther engil</td>
</tr>
<tr>
<td>inform status</td>
<td>giv</td>
<td>giv</td>
</tr>
<tr>
<td>TCS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aboutness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>definiteness</td>
<td>def</td>
<td>def</td>
</tr>
<tr>
<td>position</td>
<td>noninit</td>
<td>noninit</td>
</tr>
<tr>
<td>FBS</td>
<td></td>
<td>foc</td>
</tr>
<tr>
<td>nif</td>
<td></td>
<td>nif</td>
</tr>
<tr>
<td>cf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column 1</td>
<td>Column 2</td>
<td>Column 3</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>[I SOURCE1 (LAT)]</td>
<td>Et</td>
<td>pastens</td>
</tr>
<tr>
<td>[II SOURCE2 (AHD)]</td>
<td>uamun</td>
<td>thò</td>
</tr>
<tr>
<td>[III ALIGNMENT]</td>
<td>-L2</td>
<td>-L0</td>
</tr>
<tr>
<td>[IV POS]</td>
<td>cop</td>
<td>adv.temp</td>
</tr>
<tr>
<td>[V CAT]</td>
<td>vp</td>
<td>npe</td>
</tr>
<tr>
<td>[VI CLAUSE-STATUS]</td>
<td>main.declarative</td>
<td></td>
</tr>
<tr>
<td>[VII GF]</td>
<td>vfin</td>
<td>temp</td>
</tr>
<tr>
<td>[VIII SYL-NUMBER]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>[IX GIVENNESS]</td>
<td>now</td>
<td>giv</td>
</tr>
<tr>
<td>[X TOP/COMM]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xa ABOUTNESS]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xb POSITION]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xc TOPIC MARKER]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xd DEFINITNESS]</td>
<td>indef</td>
<td>def</td>
</tr>
<tr>
<td>[Xf FOC/BG]</td>
<td>nif</td>
<td></td>
</tr>
<tr>
<td>[Xl a FOCUS MARKER]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[XII CONTEXT]</td>
<td>Ersteinführung eines Diskursfasernten</td>
<td></td>
</tr>
<tr>
<td>[XIII COMMENT]</td>
<td>Vi gegen das Latein durch Nachstellung des Subjekts</td>
<td></td>
</tr>
<tr>
<td>[XIV BIBL]</td>
<td>20, 29 alpina no 1</td>
<td></td>
</tr>
</tbody>
</table>
Thank you for your attention

Contact
Dr. Svetlana Petrova
Humboldt-Universität Berlin
SFB Informationsstruktur, B4
s.petrova@hu-berlin.de

Project-URL
http://www2.hu-berlin.de/sprachgeschichte/forschung/informationsstruktur/index.php
http://www.sfb632.uni-potsdam.de