Information Structure and Word Order in the Early Germanic Languages and its Analysis in a Linguistic Database

Karin Donhauser, Roland Hinterhölzl, Svetlana Petrova, Michael Solf
Dept. of German Studies and Linguistics
SFB 632 / B4: "The role of information structure in the development of word order regularities in German"
Humboldt University Berlin
http://www.linguistik.hu-berlin.de/sprachgeschichte/forschung/informationsstruktur/index.php

Manfred Stede, Stefanie Dipper, Michael Götte, Julia Ritz
Dept. of Linguistics
SFB 632 / D1: "Linguistic Database for Information Structure: Formalization and Retrieval"
University of Potsdam
http://www.sfb632.uni-potsdam.de/d1

1. Research Scenario

Project B4 of SFB 632 investigates the interaction between information-structural and grammatical properties in the development of word order in the Early Germanic languages. The central issue of research is how information-structural aspects of a sentence influence its word order. In particular, we are interested in whether changes observed in the representation of information-structural categories can induce changes in the core grammar of a language or vice versa or whether such changes on both sides co-occur. The results gained from this study shed light on the complex interaction between information structure and core grammar and enable us to make important contributions to the theory of language change. By considering discourse-grammatical parameters we may be able to isolate one important contingent factor in language change and can further show how it interacts with the determinate grammatical system. We employ several tools to facilitate the analysis of information-structural properties of the Old High German language, which is the common medium of Old Germanic languages. We closely cooperate with project D1, which provides the technical infrastructure and further support for our corpus work.

2. Methods

The novelty of our approach consists in three main points:

i. variation in word order in the early Germanic languages is explained as a language-internal phenomenon going back to stylistic differences related to information packaging and discourse structure, and not as a result of language contact (cf. Fritsche 1991 and Krich/Taylor 1997).

ii. the development of methods for the analysis of information-structural phenomena in texts from historical corpora, bearing in mind that information structure is a completely new field of study in historical linguistics.

iii. investigations are carried out on a much larger scale than done in previous research on the development of Germanic syntax, including data from Old High German, Old English, Old Saxon and Old Norse.

3. Annotation & Annotation Scheme

- annotation of morpho-syntactic and information-structural features in texts from the Early Germanic period
- multiple model of information-structural representation as proposed in Möhr (1990):
  i. Pragmatic status of discourse referents: given, new, accessible
  ii. Pragmatic structure: Topic vs. Comment
  iii. Informational relevance: FOCUS vs. Background
- cumulative approach: assignment of features which are broadly associated with the information-structural categories of Topic, Pragmatic status, and FOCUS (contrast and new information)

4. Philological issues

bilingually attested translation of a Latin gospel harmony into Old High German

5. Annotation & Analysis

- For manual digitization and annotation, the partitur-editor EXMARaLDA is used, a user-friendly tool for discourse annotation with an underlying XML-format.
- Data is then converted into a generic standoff representation format, "Potsdamer Austauschstandard für Linguistische Annotationen" (PAULA), which provides import scripts for a variety of Translation tool formats (e.g. TIGER-XML, RST Tool, MMAX and EXMARALDA).
- For exploration and analysis, "ANRIS: a linguistic database for ANnotated Information Structure" - a web application accessible through standard web browsers for visualizing and querying the data is employed.
- The ANNIS Query language offers a rich set of query operators for complex multilevel annotation, including dominance and precedence relations.

6. Case Study: Information Structure and Word Order in the Old High German Tatian-Translation (9th century)

7. Example Analysis

- Task: the interaction between information structure and verb placement in Old High German
- Example: verb placement in sentences of the presentational/thetic vs. categorical distinction (Sauser 1995)

8. First Results

- presentation sentences

9. Analysis with ANNIS: Results

- Exemplary research question: What is the quantitative relation of structures i) and ii) in presentational sentences?
- ANNIS Queries:
  Q1: 01: Query for all presentational sentences
  Q2: for main clauses with presentational sentences printed in the ... (58) AND there is no adverbial in the clause
  Q3: for frame adverbial precedes the vp

10. Future Work

- Optimization of current annotation
- Development of queries and statistical motivation for several working hypotheses, e.g. the quantitative relation of patterns i) and ii) according to different scribes of the OHG Tatian
- Further annotation according to initial observations concerning the role of discourse organization (e.g. TOPIK for discourse relations) in the system of the Old High German (cf. Hinterhölzl & Petrova 2005)

References