



# 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 Germanic"  
Humboldt University Berlin  
<http://www.linguistik.hu-berlin.de/sprachgeschichte/forschung/informationsstruktur/index.php>

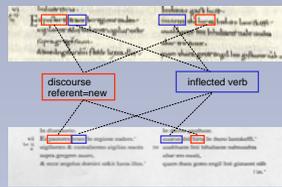
**Manfred Stede, Stefanie Dipper, Michael Götz, Julia Ritz**  
Dept. of Linguistics  
SFB 632 / D1: "Linguistic Database for Information Structure: Annotation and Retrieval"  
University of Potsdam  
<http://www.sfb632.uni-potsdam.de/d1>

## 1. Research Scenario

Project B4 of SFB 632 investigates the interaction between information-structural and grammatical conditions in the development of word order regularities in the Germanic languages. The central issue of research is how the information-structural role of a constituent influences its positioning in the clause. In particular, we are interested whether changes 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 simply 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 deterministic grammatical system. We employ corpus-linguistic methods both for creating, representing and analyzing data of Germanic languages. We closely cooperate with project D1, which provides the technical infrastructure and further support for our corpus work.

## 4. Philological issues

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



St. Gallen, Stiftsbibl. Cod. 56  
Facsimile of the manuscript  
pag. 35  
(Sonderregger 2003, p. 130)

Text-edition by Masser 1994

evidence for authentic word order is assumed in sentences deviating from the word order of the original (cf. also Dittmer/Dittmer 1998)

## 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 Austauschformat für Linguistische Annotationen" (PAULA)**, which provides import scripts for a variety of annotation tool formats (e.g. TIGER-XML, RSTTool, MMAX and EXMARALDA)
- For exploration and analysis, **"ANNIS: 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

## 9a. Analysis with ANNIS: Query formulation

Exemplary research question: **What is the quantitative relation of structures i) and ii) in presentational sentences?**

**ANNIS Queries:**

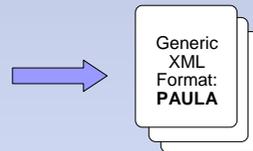
- Q1: Query for all presentational sentences**  
 clause-status=main\* & context="Ersteinführung"  
 query for main clauses with presentational sentences (marked at the context-layer)
- Q2: Query for the form: "FOC[Vfin ... DRnew ...]"**  
 (clause-status=main\* & context="Ersteinführung") AND  
 ((pos=adv\* & clause-status) OR either there is no adverbial in the clause  
 (pos=adv\* & cat=vp & #5..\* & #6)) or the adverbial does not precede the vp
- Q3: Query for the form: "Frame FOC[Vfin ... DRnew ...]"**  
 (clause-status=main\* & context="Ersteinführung") AND  
 (pos=adv\* & cat=vp & #3..\* & #4) a frame adverbial precedes the vp

## 2. Methods

The novelty of our approach consists in three main points:

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

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



## 3. Annotation & Annotation Scheme

- annotation of morpho-syntactic and information-structural features in texts from the Early Germanic period
- multi-layer model of information-structural representation as proposed in Molnár (1993):
  - Pragmatic status of discourse referents: given, new, accessible
  - Predicational structure: Topic vs. Comment
  - Informational relevance: Focus vs. Background
- cumulative approach: assignment of features which are broadly associated with the information-structural categories of Topic (referentiality, definiteness and position) and Focus (contrast and new information)

## 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. categorial distinction (Sasse 1995)

- Lat. *Vidua autem quaedam erat/ In ciuitate illa*  
OHG *uwas thar ouh sum uuitua/ In thero burgi* (T 201, 2)  
was there a widow too in that city
- Lat. *erant autem ibi lapidee hydris thar uuarun steininu uuazzarfaz* (T 81, 26)  
OHG *ther were waterpots of stone*
- Lat. *ego sum pastor bonus. bonus pastor/ animam suam dat pro ouibus suis*  
OHG *ih bin guot hirt. guot hirt/ tuot sina sela furi sinu scaph.* (T 225, 16f.)  
I am a good shepherd. [The] good shepherd gives his soul for his sheep.

## 8. First Results

- presentational sentences**
  - FOC[Vfin ... DRnew ...]
  - Frame FOC[Vfin ... DRnew ...]
- categorial sentences**  
TOP[DRgrv] FOC[Vfin ...]

- There is a correlation between information structure and verb placement in Old High German
- Vfin opens new-information focus domain
- Vfin separates an aboutness-topic from the rest of the sentence

## 9b. Analysis with ANNIS: Results

Results of querying the Tatian Corpus with 1748 documents

- Q1: 12 matches  
Q2: 6 matches  
Q3: 5 matches  
(1 match displays a different pattern)



### Preliminary conclusions

Both patterns occur nearly equally often. This provides evidence for variation within the functional domain of presentational sentences, which indicates that we encounter conditions for a language change situation in this functional domain (cf. Lightfoot 1999).

## 10. Future Work

- Optimization of current annotation**
- Development of queries and statistical motivation** for several working hypotheses, ex. 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. RST for discourse relations) in the syntax of Early Germanic (cf. Hinterhölzl & Petrova 2005)

### References

Dittmer, Arne und Dittmer, Ernst. 1998. Studien zur Wortstellung – Satzgliedstellung in der althochdeutschen Tatarüberetzung. Göttingen: Vandenhoeck & Ruprecht.  
Hinterhölzl, Roland & Petrova, Svetlana. 2006. *Rhetorical Relations and Verb Placement in the Early Germanic Languages: Evidence from the Old High German Tatian Translation (9th century)*. In Manfred Stede et al. (eds) *Salience in Discourse: Multidisciplinary Approaches to Discourse*. Münster: Salinghoff, 74–79.  
Kroch, Anthony and Taylor, Ann. 1997. *Verb Movements in Old and Middle English: Dialect Variation and Language Contact*. In *Presenting the Past: Linguistic Change and the English Language*. Cambridge: CUP.  
Lightfoot, David. 1999. *The Development of Language Acquisition, Change, and Evolution*. Blackwell.  
Molnár, Susana. 1993. *Pragmatic and Grammatical Variation and Change in Old English*. In *Information Structure*. hg. M. Reis. 156–201. Tübingen: Narreray.  
Petrova, Svetlana. 1991. *Phrase Structure in Composition: Variation and Change in Old English Word Order*. Doctoral Dissertation. University of Pennsylvania, Philadelphia.  
Sasse, Hans-Jürgen. 1995. *Theory and VS order: a case study*. *Sprachtypologie und Universalienforschung* 48, 112–3–31.  
Sonderregger, Stefan. 2003. *Althochdeutsche Sprache und Literatur. Eine Einführung in das Althochdeutsche. Darstellung und Grammatik*. 3., durchgesehene und wesentlich erweiterte Auflage. Berlin-New York: de Gruyter.