

**Methodological problems
of the information-structural analysis of data
from historical text corpora**

On the methods of information-structural analysis in historical texts: A case study on Old High German

Svetlana Petrova and Michael Solf

1. Introduction

It is an undisputable fact that information-structural research made considerable advance in the past few decades by exploring a wide range of phenomena in the systems of present-day languages spoken worldwide. Historical linguistics has also been interested in the matter of how old languages attempt to achieve the needs of communicative explicitness and stylistic expressivity. Leaving aside a huge variety of accounts within the ancient and medieval grammar tradition, even more recent descriptive works on the grammar of the ancient languages give direct or at least implicit accounts of grammatical phenomena which correlate with properties of contextual reference and pragmatics in these languages. It is most revealing that one of the earliest works that influenced the emergency of what we call 'Functional Sentence Perspective' or 'Information Structure' today, the seminal work of Henri Weil ([1844] ³1879), was dedicated to the sentence structure of ancient Greek and Latin which he regarded as a reflection of the natural flow of thoughts both in conditions of neutral as well as emphatic speech.

In more recent times, information-structural features are still a key to the proper understanding of historical text sources. Taking as an example the role of information structure for the interpretation of Biblical Hebrew texts and Old Hebrew narratives, only within a short period of time a series of monographs like those by Disse (1998), Heimerdinger (1999) and Floor (2004) appeared. As for the early Germanic languages, properties of the information-structural organization of the utterance have regularly been addressed in earlier studies aiming at explaining word order variation and verb placement. Accounts like those of Bean (1983) and Stockwell (1984) on Old English syntax have drawn a connection between the pragmatic features of sentence constituents and their placement in the structure of the clause, while in Hopper's (1979a and b) typological account, different word order patterns in Old English are directly associated with properties of discourse organization, namely with the function of distinguishing parts

of foregrounding vs. backgrounding in narratives. Similar attempts are found in the syntactic exploration of the remaining early Germanic languages as well. As early as in the accounts on the syntax of Old Saxon by Ries (1880: 1–11) or Old High German (OHG) by Behaghel (1932, IV: 3–9), among many others, the distribution of different word order patterns involves terms and concepts which from the perspective of modern linguistics clearly pertain to the field of information structure and discourse analysis.

Meanwhile, information structure has started to be taken into account as a factor responsible for the emergency of novel word order patterns and constructions and thus introducing variation as a precondition of language change in the course of language development, cf. Hinterhölzl (2004, in this volume). This view opens new perspectives on the explanation of word order variation and syntactic change in language history. Given the considerable advance in recent studies on information structure, there is a demand for a more systematic, large-scale description of the information-structural properties of sentence constituents and their syntactic realization in early Germanic.

As promising as it seems, such an enterprise is concerned with a series of methodological problems and questions which have to be addressed prior to any information-structural analysis of a historical text. The present paper aims at providing a method of dealing with information-structural questions on the basis of evidence from OHG and reflects the experience of a research project investigating the role of information structure for the development of the word order regularities in the Germanic languages.¹

Problems arising in dealing with information structure in a text from a historic corpus can broadly be classified into two major groups. The first one concerns the quality of the written data available to us, i.e. the reliability of the written text sources as well as the authenticity of the constructions and patterns provided in them. The second problem concerns the information-structural analysis proper, i.e. how to handle the huge variety of theoretical notions and terminological items used in present day's research and how to apply them to data from historic corpora.

2. The philological issue

Any research on syntax and information structure in a historical text has to address the reliability of its data base, i.e. the question whether the word order patterns and syntactic constructions provided in the written texts

sources are representative for the system of the language under investigation. In this respect, OHG displays an especially difficult ground for any conclusions on syntax, as was also recently shown by Fleischer (2006). The major part of the Old High German corpus is made up by two types of texts: i) vernacular translations of Latin biblical or other religious sources (Isidor (late 8th century), Tatian (9th century), Notker (11th century)) as well as ii) metrical poetry (the poetic Gospel Harmony of Otfrid von Weissenburg, 9th century). In both types of texts, it cannot be excluded that the ordering of sentence constituents is subject to influences imposed either by the syntax of the Latin original or by the properties of the poetic form like rhyme and metre. Therefore, we have to conclude that none of the Old High German sources of a considerable size may be viewed as a good example of authentic prose representative for the system of the dialects spoken at the period of time.

For this reason, prior to the syntactic investigation of OHG, some methodological solutions are required. Let us analyse in more detail the specific problems that the individual sources of the Old High German corpus display and consider texts that can be assumed to be more appropriate for the purpose of the intended study than others. E.g., Otfrid's Gospel harmony, the only poetic record among the more extensive ones, has been viewed to display structures and other grammatical phenomena explainable only with respect to the purposes of end rhyme and metre (cf. Ingenbleek 1880, Fleischer 2006: 35–37). Looking at the translations, there are above all two records of a considerable size: the so called Old High German Tatian and Notker's extensive work. Both expose the problem of Latin influence, and in addition to that, Notker's work is simply too young to draw conclusions from it on the earliest stages of the syntax of OHG.

The remaining text, the Old High German Tatian translation, is a record of a most disputed quality with respect to syntactic analysis, while in the fields of phonology and morphology, it has helped tremendously to shape our view on OHG. This Gospel harmony, which only survived as one single manuscript (St. Gallen Cod. 56), represents the largest Old High German text written before 850, a translation from the Latin, a bilingual source which provides the original and its translation in two juxtaposed columns. Due to the fact that a comparison of the Old High German text with the Latin source is immediately possible, an overwhelming syntactic similarity between the Latin source and its Old High German translation is visible at once. Consider the example in (1) in which the Old High German sentence follows the structure of its Latin original without any exception:

- (1) *inti quamun fugala inti frazun thiu* (T 108, 2)
 and came birds and ate them
 ‘and birds came and ate them’
 & uenerunt uolucres & comederunt ea

Given this, many philologists including Sievers (1895:LXX) classified Tatian as a “slavish imitation of the original” which is worthless for any account on syntax.

This general skepticism is furthermore fed by the fact that we discover a number of syntactic loan constructions all over the text of Tatian. This becomes particularly clear when it comes to the translation of participial constructions typical for the Latin like the *ablativus absolutus*. The scribes quite often imitate it by an Old High German *dativus absolutus*, a construction which is viewed to be a syntactic loan occurring in Old High German translations only, rather than in native writings (cf. Lühr 2005):

- (2) *Inti árleitten fon erdu skeffun / forlazenen*
 and brought_{part} to earth_{dat} ships_{dat} forsaken
allen folg&un imo (T 56, 11–12)
 all followed him
 & *subductis* a terra *nauibus* / *relictis omnibus.* secuti sunt eum;
 ‘And when they had brought their ships to land, they forsook all,
 and followed him.’

This might sound discouraging for any further work on the syntax of Tatian. However, investigations on other aspects of the language of Tatian, as the use of subject pronouns (cf. Eggenberger 1961) or the rendering of AcI-construction of the Latin original (Dentschewa 1987), show that the Old High German text does not in each case imitate the original but systematically implements different means of expression, therefore displaying some genuine features constitutive for OHG. Accordingly, Eggenberger (1961: 87) classifies the Tatian as a kind of ‘mixed form’ placing it between the type of interlinear translations like the Old High German Benedictine Rule and free translations like Isidor.

As far as word order is concerned, as early as in the investigation by Ruhfus (1897) it has been noted, that there are numerous instances in which the Old High German text deviates from the structure of the underlying original. Ruhfus explicitly bases his analysis on such cases pointing out that these are appropriate for research on syntax:

“dass eine untersuchung der wortstellung des ahd. Tatian [...] bei all seiner sklavischen abhängigkeit von der lateinischen vorlage ebenso beachtenswerte ergebnisse liefern muss, wie gleiche untersuchungen an der freien und gewandten übersetzung des Isidor oder an der dichtung Otfrids, wenn man sich nur auf eine zusammenstellung der abweichungen des ahd. vom lat. beschränkt und daraus schlüsse zieht.”

‘that an investigation of the word order of the OHG Tatian [...] despite all his slavish dependency on the Latin original ought to give remarkable results as much as similar investigations of the more free and elegant translation of Isidor or of the poetic work of Otfrid, if one only considers the collection of deviations of the OHG from the Lat. and draws conclusions from it.’ (Ruhfus 1897: 1)

In more recent times, Dittmer and Dittmer (1998) also provide a syntactic analysis of the Old High German Tatian based on instances of word order differences to the original. As a methodological pre-condition, they assume that such instances are to be taken as evidence for authentic Old High German syntax: “In the first place, only those instances which differ from the original can be taken as evidence for genuine Old High German word order.” (“Beweiskräftig für genuin althochdeutsche Wortstellung sind in erster Linie nur die von der Vorlage abweichenden Belege.”) (Dittmer and Dittmer 1998: 36).

The idea that research on word order in OHG is possible on the basis of examples showing a word order which differs with respect to the original is supported by new insights into the translation technique of this text provided by Masser (1994). He discovers that each line in the Old High German text translates exactly the same material found in the corresponding Latin line. Among thousands of lines we only find relatively few cases neglecting this principle, and the majority of them are found on the few pages contributed by scribe ϵ^2 . Astonishingly, this point seems to have passed unnoticed until recently: the former authoritative edition of Sievers (1892) at least and others did not pay the necessary attention to this. Studies on word order ignoring the role of this translation principle threaten to base their conclusions on false premises. The translation technique implemented in Tatian certainly imposes restrictions on the possibility of rendering a genuine word order pattern to the structures found in the original, cf. Masser (1997a and b). But at the same time, deviations from the word order of the Latin text within a line can be valued as evidence of genuine syntax and thus be made an object of separate investigation.

Under these conditions, the text of the Old High German Tatian may be considered as one of the most abundant data collections for research on

Old High German syntax, as recently shown by Fleischer, Hinterhölzl and Solf (2008). The fact that we can make direct comparison with the Latin part is an especially helpful device to take control over the influence of the Latin structure. These considerations motivate our decision to base our data collection and initial investigations on syntactically deviating structures of Tatian. This should not imply that we simply assume that the remaining texts are unsuitable for an account on syntax. We are aware that data from other sources is needed for comparison, and we strongly appreciate the efforts of other specialists to provide methods of dealing with these texts as well, cf. the contributions by Lühr, Lötscher and Schlachter in this volume.

In the following, we want to provide a classification of the different types of deviations encountered in Tatian according to the edition of Masser (1994).³ Altogether we find several thousands of relevant instances, an amazingly high number of significant divergences from the word order of the Latin original. We first consider the frequent changes within the boundaries of syntactic constituents and then turn to changes which surpass the constituent boundary.

2.1. Changes within the constituents

An outstandingly great number of divergences take place within the boundaries of syntactic constituents. This is the case when different types of modifiers, e.g. adjectives, possessive pronouns or genitive attributes, which are post-nominal in Latin are realized pre-nominally in OHG, cf. (3):

- (3) *uuard gifullit heilages geistes* (T 31, 9–10)
 was filled holy_{gen} spirit_{gen}
 ‘was filled by the Holy Spirit’
 Impl&us est spiritu sancto

Concerning issues of information structure and discourse organization, the regular use of the demonstrative pronoun as an article appears to be a notable feature, cf. (4):

- (4) *ther kneht uuvohs* (T 32, 6)
 the boy grew
 ‘the boy grew’
puer autem crescebat

2.2. Changes involving the order of constituents in the sentence

Changes affecting the relative order of sentence constituents are especially important for issues of word order and information structure. Here we can roughly distinguish the following cases: i) a constituent missing in the original is inserted in the translation; ii) the relative order of constituents is changed with respect to the original; iii) the translation differs from the original in providing an idiomatic, more precise expression, or it even provides supplementary information.

An important subgroup is constituted by instances of Latin synthetic forms which lack a synthetic counterpart in OHG. Although the process is confined to the verb phrase, and strictly speaking not surpassing the constituent boundary, too, the translator(s) had to dissolve the form into an analytic construction and had to decide on the placement of the finite auxiliary with respect to the non-finite verb. As examples (5) and (6) show, both orders, i.e. Aux – Participle and Participle – Aux, occur in Tatian:

(5) *thaz uuari gifullit* (T 34, 27)
 that became fulfilled
 ‘that it might be fulfilled’
 ut adimpler&ur

(6) *thiu dar giquetan ist ephrem* (T 234, 27)
 which there called is Ephrem
 ‘which is called Ephrem’
 quæ dicitur ephrem

2.2.1. Insertion

Different types of constituents lacking a lexical equivalent in the original are inserted in the vernacular texts, e.g. subject pronouns (7)⁴, object pronouns (8), adverbials (9), and finite verb forms (10):

(7) *oba ir hab& giloubon* (T 200, 12)
 if you have faith
 ‘if you were true believers’
 si habueritis fidem

- (8) *thó antlingota imo ther heilant* (T 52, 30)
 then answered him the Saviour
 ‘then the Saviour answered’
 Respondit ihesus
- (9) *Inti só sie thó ubar feritun* (T 20, 17)
 and when they there over went
 ‘and when they crossed there’
 Et cum transfr&assent
- (10) *Inti Ira namo uuas elisab&h* (T 26, 2)
 and her name was Elisabeth
 ‘and her name was Elisabeth’
 & nomen eius elisab&h

In addition, we also find instances where Latin constituents, especially particles, are not given an equivalent in the translation, cf. lat. *ergo* in (11):

- (11) *thane thú tuos elimosinam* (T 66, 29)
 when you do charity
 ‘when you do charity’
 cum facies elimosinam ergo

2.2.2. Transposition

For investigations of the conditions on word order the regular transposition of constituents are of the greatest importance. Foremost, the placement of the finite verb is of special interest in this connection. The finite verb may be shifted to the first (12), second (13) or a later position (14) within the sentence:

- (12) *Inti legitum iro hant In then heilant anan* (T 297, 20)
 and laid their hands in the Saviour on
 ‘and laid their hands on the Saviour’
 & manus iniecerunt In ihesum
- (13) *elisab&h uuas unberenti* (T 26, 7)
 Elisabeth was barren
 ‘Elisabeth was barren’
ess& elisab&h sterilis

- (14) *mitthiu thie postoli quamun / zi themo heilante*
 when these messengers came to the Saviour
 ‘and when the apostles came to the Saviour’
 Et cum uenissent apostoli / ad ihesum (T 102, 29–30)

Good evidence for authentic structures is especially given in the following two cases. First, multiple changes may indicate that the translator paid more attention to the stylistic quality of his translation. The more is changed, the more we ought to assume, that an authentic structure is expressed. Cf. (15) where the transposition of the copula *sint* is being accompanied by the insertion of the subject pronoun *sie*:

- (15) *sie sint blinte inti blintero leitidon* (T 128,14)
 they are blind and blind-GenPl leaders
 ‘They are blind and leaders of the blind’
 caeci sunt duces cecorum

The second group contains the cases of transpositions violating the line boundary:

- (16) *ih ni / haben diuual* (T 219, 8–9)
 I Neg-have devil
 ‘I don’t have the devil’
 ego demonium / non habeo

The authenticity of the transposition in this example is being confirmed by the fact that the scribe neglects the line boundary, which strongly violates the main translational principle of this text. Otherwise, whereas this example *prima facie* indeed points at a certain freedom in translating the Latin, there are other cases where the violation of the line boundary is not an intended means: it might have been for reasons of space or simple carelessness⁵ as well. The mere consistency of the translation with respect to the line-for-line principle might lead to the conclusion that every single case of its violation has to be of particular value. However, this is not always the case, for instance, when the Latin word is divided by the line boundary and the scribe was forced to decide where to chain up the translation:

- (17) *Intj / uûizagon sint ouh tote* (T 219, 23–24)
 and prophets are also dead
 ‘and the prophets are dead as well’
 & pro / ph&tæ mortui sunt

A higher degree of certainty is only possible by comparing similar instances.

2.2.3. Free translations

Considering the limited freedom of the translation some regularly occurring instances are especially remarkable which show constructions with a sense more or less diverging from the original. We may assume, that the Old High German expression grants a better understanding and is therefore preferred to a word for word translation. As evidence for authentic OHG they are of a particular high value.

Example (18) shows, that the range of changes may go as far as to forming of a relative clause without a Latin counterpart:

- (18) *fon theru burg thiu hiez nazar&h* (T 35, 15)
of the city which is-named Nazareth
‘out of the city, which was called Nazareth’
de ciuitate nazar&h

In (19), the subject *gotes engil* is related to a new finite verb *quam*, which the translator preferred to the possible literal equivalent *inti senonu engil gotes gistuont nah in*:

- (19) *quam thara gotes engil Inti gistuont nah in*
came there gods angel and stood near him
‘there the angel of god came and stood near them’
& ecce angelus domini st&it Iuxta illos (T 35, 32)

Finally, example (20) demonstrates a Latin participle construction formed into a temporal subordinate clause:

- (20) *thô thaz gihorta herodes ther cuning* (T 39, 17)
when that heard Herod the king
‘when Herod the king heard that’
audiens autem herodes rex

2.3. Some doubtful cases

The different instances of divergences require an interpretation within the respective context. It is sure, for example, that single divergences do not always guarantee the authenticity of the whole sentence. Example (21) provides a transposition of constituents within the absolute construction. As we stated above, this kind of construction represents a syntactic loan, and therefore it is misleading to assume an authentic Old High German example in this case:

- (21) *Inti Inphanganemo antuuvrte In troume* (T 40, 20)
 and received_{dat} answer_{dat} in dream
 ‘and having received the answer in a dream’
 & responso accepto In somnis

The structure in (22) provides us with another interesting example of the difficult approach to Latin participle constructions. Despite of the transposition of the object pronoun and the dissolution of the Latin passive infinitive the translation here follows the original *accusativum cum infinitivo* by risking to make the whole sentence incomprehensible:

- (22) *uuenan her uuolti Inan genemnitā uuesan* (T 30, 29)
 who_{akk} he wanted_{subj} he_{akk} named_{akk} be
 ‘*whom he would him have called’ / ‘how he would have him called’
 quem uell& uocari eum

As a source for Old High German syntax this example is at least of a doubtful quality. It mixes up features of a different origin: while the subject pronoun is being inserted and the object pronoun is being transposed, oddly enough the accusative of the interrogative pronoun is being kept, and the *accusativum cum infinitivo* that is maintained transpires the character of a Latin sentence.

To conclude, superficial decisions about the authenticity of syntactic structures in Old High German sources risk to ignore the multiple problems connected with the way the corpus is handed down to us. As has been demonstrated for the case of the Old High German Tatian, the possibility of comparing the translation with its original definitively supports the search for authentic syntactic structures.

3. The theoretical issue

The second type of methodological problems in need of clarification concerns the information-structural analysis in texts from historical corpora. This question itself involves two aspects to be discussed separately. The first one concerns the conceptual and terminological determination of the information-structural categories. As information structure involves some of the most controversial issues in today's linguistic research, we have to cope with a number of different notions and terminological expressions for the main categories like topic and focus. Second, we have to look for a way of identifying these categories in texts which are available in written form only, thus offering little clues to the prosodic realization of the utterance.

3.1. The definition of information-structural categories

Most researchers agree on the notion that information structure concerns the formal realization of an utterance according to the context as well as to the state of informedness of the interlocutors (cf. Lambrecht 1994: 6). It is also well known that information structure manifests itself in a variety of domains, such as phonology, morphology and syntax, though the utilization of formal means marking information-structural properties strongly differs from language to language.

On a par with the great diversity of formal means used to express information-structural properties on different levels of language structure, current research on information structure is characterized by a great number of theoretical approaches to the explanation of the central notions in question which has led to a considerable amount of conceptual and terminological divergence and confusion, cf. Musan (2002: 202–208). Literally every notion or feature that has been proposed as relevant to information structure is subject to debate and exhaustive discussion in the literature, and there is no agreement on the definition of any category that has been proposed to capture information-structural properties in language.

In fact, this situation points at the immense complexity of information structure as a linguistic phenomenon. An approach that promises to yield good results in capturing all relevant aspects of information packaging should therefore aim at isolating different functional dimensions of information structure into which a more precisely defined inventory of categories may be applied. Such an approach was proposed by Molnár (1993) and adopted for the information-structural analysis of both modern and extinct

languages, e.g. by Zybatow (1999) for Russian or by Disse (1998) for Biblical Hebrew. Adopting this approach, we assume that there are at least three functional layers on which properties related to information structure have to be distinguished:

- i. the pragmatic status of sentence constituents, i.e. the distinction between **given** vs. **new** information in discourse as seen in the classical opposition of theme vs. rheme
- ii. the predicational structure of the utterance allowing for a bipartite division of the sentence into a starting point, or **topic** of the utterance and a predication, or **comment** on this topic
- iii. the distinction of **focus** vs. **background** in terms of communicative weight or relevance for the development of the discourse.

These layers of information structure are viewed to function independently in language but to interact with one another thus yielding the full picture of the information-structural shaping of an utterance. Decomposing information structure and creating a multi-layered model in the way proposed by Molnár (1993) and Krifka (2007) bears substantial consequences for the understanding of the information-structural categories and their proper description in theory. First, and contrary to a common assumption in previous literature, topic and focus need not to be mutually exclusive but are allowed to overlap as they pertain to different layers of information-structural segmentation. Furthermore, a model like this offers a way to avoid the common practice of identifying the topic of an utterance with the contextually given information as well as of restricting focus to new information only.

In this sense, the segmental approach provides the basic scheme for a more fine-grained analysis of the information-structural value of individual parts of an utterance. It allows for the assignment of a broad variety of information-structural features as well as for the evaluation of how different combinations of features influence the positioning of constituents in the clause. As shall become clear later, such a cumulative approach is of inestimable value for the information-structural analysis of sentence constituents in text from historical corpora. Nevertheless, the explanation and the catalogue of categories that make up the functional dimensions of information structure in the model given above are far from being unproblematic and unquestionable. The following elaboration gives an overview over the most relevant concepts concerning the information-structural categories as well as over some main points in the discussion on them in current research.

3.1.1. *Given vs. New*

Starting with the distinction between given vs. new information in a sentence, it may appear that the assignment of these categories is the most unproblematic one in analysing information structure in a running text of the historic corpus of a language. However, the discussion provided in Prince (1981: 225–232), among many others, points at the different aspects these notions have acquired in literature as well as at some considerable problems in applying them to natural data:

- given vs. new in the sense of **shared knowledge**: given is the information that the speaker believes the listener already knows and accepts as true, whereas new is that information which the speaker believes that the listener does not yet know, cf. Clark and Haviland (1977: 4)
- given vs. new in the sense of **cognitive activation/salience**: “Given (or old) information is that knowledge that the speaker assumes to be in the consciousness of the addressee at the time of the utterance. So-called new information is what the speaker assumes he is introducing into the addressee’s consciousness by what he says.”, cf. Chafe (1976: 30)
- given vs. new in the sense of **predictability/recoverability**: “An element in a sentence represents old, predictable information if it is recoverable from the preceding context; if it is not recoverable, it represents new, unpredictable information.”, cf. Kuno (1978: 282–283)

Prince (1981) shows that the definitions of ‘given’ vs. ‘new’ proposed in the literature – though being not completely independent – bear significant differences in various important points. In comparing the concept of Clark and Haviland (1977) with the one of Chafe (1976) it becomes obvious that both notions explain the given/new-distinction in terms of cognition since both ‘activation’ and ‘knowledge’ concern the cognitive state of the interlocutors. Nevertheless, a basic difference appears: an entity that is not explicitly pre-mentioned but inferable from the common knowledge of the interlocutors counts as given in the sense of Clark and Haviland (1977) but as new, i.e. known but not necessarily activated, in the concept of Chafe (1976). Accordingly, this yields two different interpretations of one and the same constituent (*your father*) in a sentence like (23)⁶:

- (23) I saw *your father* yesterday.
- a. I saw *your father*_{giv} yesterday
given=known, i.e. inferable according to Clark and Haviland (1977)
 - b. I saw *your father*_{new} yesterday
new=not activated according to Chafe (1976)

At the same time, as Prince (1981: 229) also remarks, the expression *a two-headed man* in (24) should gain – according to Chafe’s definition – the same status as the expression *your father* in (23b), a situation that is clearly counter-intuitive:

- (24) I saw *a two-headed man*_{new} yesterday
new=not activated according to Chafe (1976)

Another problem arises with respect to the uniformity of formal means expressing givenness in language. According to Clark and Haviland (1977) the referent of the definite expression *the beer* both in (25a) and (26a) is given by virtue of its status as being known via contextual pre-establishment in (25) and inferentiality in (26), respectively. Nevertheless, the ‘given’ referent can be pronominalized only in the case of (25b), cf. the ungrammaticality in (26b):

- (25) a. We got some beer out of the trunk. *The beer* was warm.
b. We got some beer out of the trunk. *It* was warm.
- (26) a. We got some picnic supplies out of the trunk. *The beer* was warm.
b. We got some picnic supplies out of the trunk. **It* was warm.

Following this argumentation, Prince arrives at the conclusion that the traditional dichotomy of ‘given’ vs. ‘new’ is evidently too narrow to capture the fine but nevertheless significant differences regarding the activation state of referential expressions in natural discourse. As a result, she proposes to suspend the old dichotomy and exchange it for a model of a scalar representation of features in which ‘given’ and ‘new’ are only the two endpoints in a wide range of subcategories.

Prince’s proposal involves a triple of notions under the core idea of “assumed familiarity”: i) new, ii) inferable and iii) evoked. Each of these pragmatic states is subdivided into deeper categories. So i) the notion of

'newness' falls into the subcategories of i-i) 'brand-new' items whose existence in the mental world of the listener is being created at the time of the utterance, and i-ii) 'unused' items which are 'known' to the hearer but not activated at the time of utterance. The introduction of brand-new referents is itself divided into i-i') 'anchored' vs. i-i'') 'unanchored' due to the fact that a brand-new item may be established in the context as linked to another entity – an anchor – that is not brand-new itself. This occurs e.g. in the case of modified NPs like *a person I know* or *a girl I work with*. On the other pole of the scale, Prince considers iii) the category of entities already 'evoked' in the discourse model. These categories are either iii-i) evoked textually, i.e. they are explicitly mentioned in previous context, or they are iii-ii) evoked situationally, e.g. when they refer to the interlocutors themselves or to referents which are salient in the communicative situation. In between, Prince distinguishes an additional, and more complex, category of ii) 'inferable' referents whose mentioning in an utterance may logically be inferred by the listener according to the entities already evoked in the context. Here, one subclass is represented by the so-called ii-i) 'containing inferables', i.e. by referents staying in a set-member relationship (*picnic supplies-beer*), and another one is formed by the so-called ii-ii) 'non-containing inferables', i.e. by referents staying in an analogy relation to one another (*bus-driver; party-music* etc.).

Building upon this new taxonomy of the informational state of referents, Lambrecht (1994: 74–113) accounts for some formal means typically correlating with them: i) presence vs. absence of accent; ii) pronominal vs. lexical coding; iii) definite vs. indefinite marking. According to the different kinds of combination of features in the actual realization of referents of these pragmatic classes, Lambrecht maps these categories onto an ordered relation, a so-called scale of identifiability of referents presented with some supplementary terminology in (27), cf. Lambrecht (1994: 109):

- (27) i) unidentifiable/brand-new
 ii) unidentifiable anchored/brand-new anchored
 iii) inactive/unused
 iv) textually accessible
 v) situationally accessible
 vi) inferentially accessible
 vii) active/given

This scale of identifiability correlates with the formal explicitness of referents in discourse. The lower a referent is to be classified according to the

scale in (27), the less explicit its formal representation is in real communication. Turned the other way round, the growth of formal explicitness in the representation of a referent points at a higher need to activate or re-activate the identification state of the referent.

Furthermore, Lambrecht (1994: 113–115 and 165) refers to a substantial cross-relation between the identification state of a referent and its appropriateness to provide the topic of an utterance. He argues that the lower a referent is mapped on the scale of identification, the more it is likely to act as the topic of an utterance. In doing so, Lambrecht tangles one of the most problematic and difficult questions in information-structural research, the relationship between givenness and topicality in language, a question to be discussed in the following section.

3.1.2. Topic vs. comment

According to the multi-layer approach on information structure proposed by Molnár (1993) and advocated in section 3.1 above, the topic vs. comment distinction refers to the possibility to identify a bipartite division of the predication in terms of an item used as the starting point, or topic and another one providing the predication, i.e. the comment on this topic. A classical form of this kind of predicational separation occurs in copular constructions or generic utterances of the kind given in (28) and (29) for which the basic definition provided by Hockett (1958: 201) applies: “the speaker announces a topic and then says something about it”:

(28) [The house]_T [is green.]_C

(29) [Trees]_T [have green leaves.]_C

Apart from this very basic notion originally intended to describe the constituent structure of predicative constructions, a range of competing concepts have been proposed making topic one of the most problematic terms to work with in information-structural analysis.

First of all, different syntactic strategies have been discussed as a means of marking the topic of the sentence. Constructions like Left dislocation (Frey 2005), Hanging topic or ‘As for...’ are among these. The Prague School of Functional Sentence Perspective applied a more general view by assuming the topic to be the first element in a sentence (Firbas 1966). This notion was also adopted by Halliday (1967) to cover all sorts of ‘topical-

ized' elements in a clause. For German, moving a constituent to the 'Vorfeld', i.e. to the single position before the verb in root clauses has also been discussed with respect to topicality. And although this type of dislocation has been shown to apply for other information-structural categories as well (Axel in this volume), studies like Molnár (1993) observe a strong tendency of topical material to appear clause-initially in German.

Arguments against this purely configurational account on topics come from alternative approaches to the explanation of this notion, e.g. from the aboutness-concept of Reinhart (1981) which is one of the most widely accepted ones in the literature today. Reinhart (1981) adopts the basic notion of pragmatic 'aboutness' as a distinctive feature of topichood and develops tests triggering typical topic readings. According to these, the item identified as the topic of the sentence acts as the referent *X* to be inserted in a proposition of the kind *A says about X that X...or as for X, X is...* Analysing the appropriateness of utterances gained from the insertion of different sorts of NPs in these tests, Reinhart shows that e.g. quantified NPs "are often hard, and sometimes impossible, to interpret as topics" (1981: 65). This behaviour she explains by virtue of the fact that the quantified NPs, failing to pass the topic tests, do not allow for an interpretation as referents. This is in clear contrast to universally quantified NPs denoting sets of referents and being available to act as topics. According to this, Reinhart restricts topichood to referentiality, and compares sentence topics with the referents acting as the entries of a subject catalogue in a library under which the propositions made about them are stored. Topics are in the sense of Reinhart "referential entries under which we classify propositions" (1981: 80). In this way, a sentence topic is allowed to appear anywhere in the sentence. Or put in different terms, a sentence-initial constituent which is not referential fails to classify as a topic.

Another widely accepted account on topichood is the one proposed by Gundel (1988) who relates pragmatic 'aboutness' to the conditions of givenness and accessibility. Departing from a definition of the topic-comment structure as presented in (30):

- (30) Topic definition:
An entity, *E*, is the topic of a sentence, *S*, if in using *S* the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to *E*. (1988: 210)

Comment definition:

A predication, P, is the comment of a sentence, S, if in using S the speaker intends P to be assessed relative to the topic of S.

Gundel arrives at the conclusion that in terms of felicitous communication, shared familiarity on the topic expression appears to be a necessary precondition of topichood: „An entity E can [...] serve as a topic [...] if, both speaker and addressee have previous knowledge of or familiarity with E“ (1988: 212). According to this definition, new referents are banned from acting as sentence topics, an option held open by Molnár (1993). Furthermore, the concept of Gundel aims at identifying givenness and accessibility with topichood though no identification of newness with comment is implied; cf. Gundel (1988: 212). This simply means that given referents are equally allowed to be part of the topic or the comment of a sentence, while no operational methods are proposed to decide which of more than one given or accessible referents in a clause should be taken as the topic of the sentence.

Still another account on topic is promoted by Chafe (1976) who argues for a basic differentiation between topics in topic-prominent languages like Mandarin and topics in languages like English. For the latter, he criticizes the definition of Halliday (1967) by pointing to apparent contrastive effects paired with the function of topicalization in English in sentences of the type *This play John saw yesterday* (Chafe 1976: 49). Consequently, Chafe proposes to suspend the term topic for such constructions in English. Instead, he pays more attention to the first type of topics, i.e. the so-called Chinese-style topic which he assumes to represent the prototypical case of topichood. However, analysing the interpretation of sentences containing typical Chinese-style topics, as the one in (31):

- (31) nèi-xie shùmu shù-shēn dà.
 those tree tree-trunk big (Chafe 1976: 50)

Chafe concludes that the role of the topic is not so much to say what the sentence is about, but to “limit the applicability of the main predication to a certain restricted domain” (Chafe 1976: 50). As not only nominal referents but also adverbials of time and location are allowed in this function, Chafe extends the notion of topic to the overall concept of ‘frame-setting’ used to subsume all sorts of elements providing the “spatial, temporal, or individual framework within which the main predication holds“ (1976, 50).

Comparing Chafe's notion on topic as a set of different frame-setting elements with the aboutness-concept of Reinhart (1981) and the familiarity-concepts of Gundel (1988) described above, the question arises whether both frame-setting topics and aboutness/familiarity-topic represent different parts of one and the same phenomenon or whether these are completely different phenomena to be held apart in information-structural analysis. In current research, the former option seems to prevail. Analysing different aspects of topic-comment constructions in German, Jacobs (2001) promotes the view that frame-setting and 'usual' topics might represent "two different but similar prototypes of T[opic]C[omment]" (2001: 658) thus making topicality a polysemous category. In line with this view, Maienborn et al.⁷ explore aboutness-topics and frame-setting as two different notions of sentence topichood which share some important syntactic, semantic and pragmatic properties with respect to the rest of the sentence as well as to the larger discourse.

3.1.3. *Focus vs. background*

The third functional layer within the model of information packaging, the layer of the focus vs. background distinction, is subject to no less controversy than the other two domains discussed above. There is some common agreement on the explanation of focus in terms of informational relevance and emphasis. This consideration implies two main aspects, one related to the contents of the utterance and one related to its formal realization. According to the first one, it is intuitively appealing to suggest that not all parts of the information exchanged between the interlocutors are equally important or relevant with respect to achieving the communicative goals of the utterance. Secondly, the notion of focus in terms of emphasis relates to the formal prominence of certain parts over the rest of the utterance, most commonly manifested in terms of prosodic prominence, e.g. main stress, phrasing and intonation in so-called intonational languages.

According to many standard assumptions, the informational weight of focused constituents is explained in terms of the distinction between presupposed, or given vs. newly added, or asserted information in discourse (cf. the overview in Szendrői 2004: 230). The difference between the informational relevance of parts of the utterance is best illustrated by question-answer pairs where the newly added, asserted information may be precisely identified as that part of the answer which corresponds to the wh-phrase in the preceding question. On the formal side, the newly added in-

formation is marked by means of prosodic prominence illustrated by capitalizing the word carrying sentence stress, cf. (32):

- (32) Q: What did John eat?
A: John ate PIZZA. (cf. Szendrői 2004: 230)

According to the high operational value of question-answer pairs for focus detection, most focus theories stick to exploring this phenomenon on the basis of such discourse examples only. However, it is misleading to identify focus with ‘newness’ in terms of the informational status of the referents. As many examples in previous literature have shown, information provided in the answer to a preceding *wh*-question prosodically behaves the same way as the new one in (32) even if it refers to entities which are contextually given or inferable, cf. (33):

- (33) Q: Who did Felix_i praise?
A: a. He_i praised HIMSELF_i.
b. He_i praised YOU.
c. He_i praised his BROTHER. (cf. Lambrecht 1994: 260)

With respect to this property, already Halliday (1967) warns against identifying focus with pragmatic newness: “What is focal is “new” information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse (Halliday 1967: 204).

The same kind of correlation between focus and given information regularly emerges in cases of contrastive focus. Here, a part of the utterance gains informational relevance over the rest of the sentence by virtue of the fact that it is placed in a relation of contrast to another item which is either explicitly given or implied as part of the context. Consider (34) in which the answer picks up referents selected from a set of alternatives explicitly mentioned in the question:

- (34) Q: Who did she invite to her birthday party, John_i or Peter_j?
A: She invited JOHN_i. PETER_j she cannot stand.

Lambrecht (1994: 209–218) provides some more considerations showing that the identification of focus with that portion of the answer which con-

veys the ‘new information’ is too simplifying and inaccurate. In his analysis of the example in (35):

- (35) Q: Where did you go last night?
A: I went to the MOVIES. (cf. Lambrecht 1994: 209)

he argues that the new information in the answer is neither identical with the referent of the expression *the movies* nor to be restricted to any missing referent actually. What is new is in fact the relationship between the denotatum of the expression *the movies* and the proposition made explicit in an answer of the kind in (35’):

- (35’) the place where I went last night was the movies.

Following this argumentation, focus is the device marking the establishment of a new kind of relation between a denotatum and the rest of the proposition, thus creating a “new state of information in the addressee’s mind” cf. Lambrecht (1994: 210). In this sense, a focus relation is guaranteed even in sentences operating with given material only, as is the case in (36c) of the famous syllogism or in (34) above:

- (36) a. Every human is mortal.
b. Socrates is a human.
c. Socrates is mortal.

Another approach avoiding the definition of focus in terms of newness is found in the model of Rooths (1985). His meanwhile broadly accepted method involves a definition of focus in which the assignment of a formal focus feature on a constituent signals that it has been chosen from a set of alternatives and that is inserted in an open proposition.

3.2. Access to prosodic information in OHG?

So far, the problems of the overall definition of information-structural categories have been addressed. Relating this issue to the information-structural analysis of data from the historic corpus of a language, still more difficulties arise. One major problem comes from the fact that we have to deal with a quantitatively restricted amount of data. Accordingly, no additional data elicitation is possible in this case, i.e. no tests for checking the

appropriateness of sentences as well as no negative evidence is available. This simply means that the information-structural analysis of corpus data has to do without some basic methods most widely applied in synchronic research on information structure.

Another problem arises from the fact that the data to be analysed here is attested in written form only and thus provides no reliable access to prosodic information which is crucial to the detection of a number of phenomena related to information packaging as accent/de-accenting, phrasing and intonation.

The previous literature gives sporadic accounts on indications drawn from the graphical representation of the manuscripts, though in sum, these basically concern the macro-structural dimension of text organization and episode subdivision, and less the level of prosody. For instance, we have quite reliable access to the separation of chapters (so-called 'fits') in some texts of the early Germanic corpus since these are either marked by Roman numbers as in the manuscripts of the Old English *Beowulf* and the Old Saxon *Heliand*, cf. Bästlein (1991), or introduced by individual chapter titles as in the Old High German *Gospel Harmony* by Otfrid. Simmler (1998) also shows that in the Old High German Tatian translation, numerous concordance notes attached to the Latin (and sometimes to the Old High German) part of the text give evidence on text organization and episode-division, next to the use of capital letters at the beginning of new episodes.

Such clues are not available on the micro-structural level. In fact, the most representative of the Old High German manuscripts contain a number of accent diacritics. According to recent observations, certain information on prosody can be inferred in the texts of Otfrid and Notker, cf. Kleiber (2004: 119–142) and Fleischer (in this volume). However, the functional explanation of accent placement in Tatian is more than questionable. Analysing the numerous accent diacritics found here, Fleischer (in this volume) arrives at the conclusion that these may be assigned different functions which strongly interfere with each other thus leaving little ground for any substantial conclusions on prosodic prominence.

Fleischer (in this volume) also points at spacing between words as another potential clue to prosodic information in Old High German manuscripts. The absence of a blank space between words may be interpreted as a reflection of the fact that these words were pronounced together, without a phonological break between them. Consider the different representation of the personal pronoun in (37):

- (37) *eno thu bistu mera / unsaremo fater iacobe*
 verily you are-you more our_{dat} father Jacob
 ‘And you, are you more than our father Jacob?’
numquid tu maior es / patre nostro iacob. (T 131, 18–19)

The graphical representation of cases like the second occurrence of the personal pronoun in (37) allows for the conclusion that a functional word is de-accented and cliticized to a main word as well as that no prosodic phrase boundary intervenes between these words. On the contrary, large spaces between words have been analysed as signalling a speech pause or a phrasing boundary, cf. also Masser (1997: 58).

However, this is a catalogue of potential indications whose relation to prosody cannot be claimed without any doubt. Moreover, the proper evaluation and systematic analysis of this scattered evidence on the text-organization and the prosodic realization in corpus data depends on the availability of precise diplomatic editions reflecting these properties of the graphical representation of the texts. In numerous cases, a comparison with or an exploration of the manuscripts is necessary.

These problems clearly point at the fact that for the proper information-structural analysis of sentences from historical corpora, context interpretation and discourse organization are the most reliable source to clues for the information-structural value of sentence constituents. These aspects shall be taken as a basis for a model of information-structural analysis on sentences from the Old High German Tatian proposed in the following section.

4. A proposal for the information-structural analysis of corpus data

4.1. The main scheme

Due to the lack of a proper definition of the information-structural categories which can be applied to the data immediately, and given the difficulties arising in texts of the historic corpus of a language, we propose a novel approach based on the multi-layered model on information structure presented in section 3.1 above. We maintain the notion that informational status, topic-comment and focus-background are basic distinctions which interact with one another but nevertheless reflect different aspects of information packaging. Therefore, the attributes belonging to these different distinctions have to be kept apart in real data analysis. However, we dispense with the idea of assigning the main categories of topic and focus

directly; rather, we aim at assigning to the sentence constituents a possibly wide range of features which have been connected with information structures in the previous literature. In doing so, we are able to draw conclusions about the interdependence between certain features or combinations of features on individual constituents and their positioning in the clause.

Departing from the analysis of the basic notions in information-structural research, we distinguish the following attributes and features considered as constitutive for the main information-structural categories.

4.1.1. Informational status of discourse referents: given, new and accessible

Given the argumentation in Prince (1981) discussed in section 3.1.1 above and some further considerations in Dik (1989: 268–270) and Lambrecht (1994), we agree on a scalar representation of features attested to the informational status of discourse referents. The notions ‘given’ and ‘new’ are viewed as the endpoints of the scale. However, for the sake of unambiguous applicability, these two notions are bound to textual pre-establishment: ‘given’ is restricted to expressions referring to explicitly pre-mentioned referents, while ‘new’ covers referents introduced to context for the first time.

The anaphoric expressions that serve as linguistic correlates of givenness can be represented on a scale of explicitness starting with zero-anaphors, personal and demonstrative pronouns and ending up with full expressions like full NPs, proper names and epithets. Consider the following text-example adapted from Dik (1989: 271):

- (38) Yesterday I got a phone call from *the tax inspector*_i. *He*/~~*the man*~~/
*the jockey*_i wanted me to come to *his*_i office and *he*_i / \emptyset _i gave me the impression that I was in for some trouble.

Contrary to that, the term ‘new’ covers notions that are not explicitly pre-established in previous context. Typically, a new discourse referent is introduced explicitly following one of the strategies outlined by Dik (1989: 268):

- meta-linguistic information: *I’m going to tell you a story about X*
- as an object or second argument in sentences with transitive verbs:
In ... , we saw X

- as the subject of an existential or presentation construction of the type *Once upon a time, there lived/was ...* which is typical for text-opening sentences
- with verbs of motion, denoting “appearing on the scene” (Dik 1989:268): *Suddenly, right before our eyes, X appeared ...*

Between the states of ‘given’ and ‘new’ as polar values, a whole range of categories is situated sharing the property of not being previously mentioned but nevertheless staying in a certain relation of relevance either to the communicative situation as a whole or to entities already established in discourse. These categories are subsumed under the label of ‘accessible’ entities due to the fact that they need not be introduced explicitly but are semi-active at the time of the utterance and are thus available for proper reference. We distinguish the following cases of accessibility:

- expressions referring to the **interlocutors**, e.g. the deictic pronouns of the first and second person singular/plural as well as full NP, proper names etc. that are coreferential with them
- **anchoring**: entities that are introduced in a certain relation to already activated referents; *a friend of mine ...*, *a person I work with ...*
- **bridging**: entities which stay in part-whole relationship or in a relation of analogy to an already pre-established referent; *John gave a party last week, but the music was awful ...*
- **shared familiarity**: entities belonging to the common knowledge of the interlocutors like real names, proper names, etc.

4.1.2. Predicational separation: topic vs. comment

Bearing in mind the different accounts on topicality and the lack of a commonly accepted definition of this term, we decompose this notion with respect to all features – both conceptual and formal ones – viewed to be constitutive for topicality in language:

- givenness/accessibility of referents: to be acquired from the layer of informational status, see above
- aboutness: whether or not any constituent of the sentence fits to replace X in the topic-test *A says about X that X ...*
- definiteness: whether the referential expressions involve a definite or indefinite reading

- syntactic realization in the clause: whether any of the constituents in the sentence is involved in a syntactic construction associated with topicality (e.g. Left dislocation LD, hanging topic HD, ‘As for ...’, ‘Vorfeld’ etc.)

The co-occurrence of all features on one and the same constituent would yield an optimal topic candidate; the less features apply on a constituent the less it fits to be identified as the topic of the sentence.

This method facilitates a clear decision on the assignment of the topic category in examples like the following small discourse:

- (39) a. *senonu tho uuas man In hierusalem.* / b. *thes namo*
 behold there was man in Jerusalem this_{gen} name
uuas giheizzan simeon, / c. Inti ther man uuas reht Inti gotforht.
 was named Simon and this man was just and devout
 ‘a. And behold, there was a man in Jerusalem b. whose name was
 Simeon, c. and this man was just and devout’ (Luke 2, 25)
 a. & ecce homo erat In hierusalem.’ / b. cui nomen simeon, / c. &
 homo iste iustus & timoratus.’ (T 37, 23–25)

Here, the expressions *thes namo* ‘whose name’ in (39b) and *ther man* ‘this man’ in (39c) refer back to the discourse referent *man* ‘a man’ introduced in the initial conjunct (39a). The referent *namo* in (39b) is an instance of anchoring since it is introduced in relation to an already mentioned entity, and *ther man* in (39c) is a full anaphor to ‘a man’. Both anaphoric expressions are referential, definite and realized clause-initially in the corresponding conjuncts. Therefore, these constituents can be identified as topics, while the remaining part of the utterances provides the comment on these topics, respectively. This interpretation satisfies the intuition that the sentences in (39b–c) predicate on a given entity by saying something about it, e.g. by assigning a property to it. The first conjunct (39a), however, is a case of a presentational construction establishing the topic of the subsequent utterances but displaying no topic-comment structure itself (cf. Sasse 1995: 4–5). Moreover, the referent ‘a man’ in (39a) doesn’t bear any significant topic features: it is referential but new, indefinite and occurs in a late position in the clause. For these reasons, no topic assignment applies to ‘a man’ in (39a).

However, the method described above does not allow for an unambiguous solution in each case. Consider instances like (40) in which the criteria

announced above equally apply to more than one discourse referent in the clause:

- (40) *quad Iru ther engil* (T 28, 3–15)
 said her_{dat} this angel
 ‘Then the angel said to her’ (Luke 1, 30)
 & *ait angelus ei*

The sentence is nested within a dialogue sequence in which the expressions *Iru* and *the angel* relate to explicitly pre-established referents whose relative order in the sentence is shifted in OHG against the Latin original. Although crucial criteria for topicality like givenness and definiteness apply for both of the referents involved, it is difficult to decide which of them qualifies as the aboutness-topic of the utterance. Broadly speaking, the sentence is about both of them, though in fact the application of the topic-test *A says about X that X ...* does not yield a proper paraphrase of the contents of the sentence. A more appealing interpretation is the one considering the sentence to predicate on the event itself, i.e. on the fact that the conversation between the referents is carried further. In this case, the given referents are involved in a new state of affairs but are not the topic of the clause.

Consequently, the most decisive issue in assigning the category topic to a constituent bearing any of the features relevant for topics involves the presence of a topic-comment division in the utterance: only if the sentence allows for a categorical reading (Sasse 1995: 4–5), i.e. if it announces a unit as the starting point of an utterance and then makes a statement on this unit, is it reasonable to assign a topic category to any constituent bearing any of the features relevant for topics.

4.1.3. Informational relevance: focus vs. background

In section 3.1.3 we argued for a pragmatically rather than prosodically defined notion of focus. Following the discussion in current research, we defined focus as that part of an utterance that provides the most relevant information in a particular context as opposed to the (not so relevant) rest of information making up the background. There are at least two ways for a part of an utterance to gain information-structural relevance over the rest of the sentence:

- it provides ‘new’ information in the sense that it is either requested in a preceding wh-question or necessary to develop the discourse; or
- it stands in a relation of contrast to another constituent in the discourse.

According to this, we distinguish the types of new-information focus (nif) and contrastive focus (cf), respectively. We also assume that new-information focus and contrastive focus are not mutually exclusive but may apply within one and the same domain. As we also referred to in the above discussion, the term ‘new-information focus’ is somewhat misleading as it should not equate focus with ‘newness’ but stand for the new type of relation that emerges when an open proposition is saturated by inserting the missing material out of a set of possible alternatives, regardless of the fact whether this material is known, given, inferable, etc. to the interlocutors. Related to referents, ‘new’ is only a category on the level of informational status and not a constitutive feature of focus. The assignment of the category focus to a particular part of the proposition points to the material which promotes the establishment of this new relation between referents in the addressee’s mind.

In this respect, the literature refers to the ‘focus domain’ of a sentence, i.e. to the different extension of the focus material in the utterance. Typically, the following cases are distinguished:

- the focus domain spreads over the entire proposition, e.g. in all-new sentences opening a text discourse in so-called thetic sentences as the one in (41), cf. Sasse (1995: 4–7):

(41) Q: What’s that smell?
 A: [The KITCHEN is burning.]_F

- the focus domain comprises the whole VP (42a), a single constituent within the VP (42b), or a part of a constituent (42c) respectively. The following examples show a different assignment of the focus domain for one and the same structure depending on the preceding context:

(42)	a.	Q: What is Peter doing? A: He [is reading a PAPER.] _F	VP-focus
	b.	Q: What is Peter reading? A: He is reading [a PAPER.] _F	XP-focus

- c. Q: What sort of paper is Peter reading?
 A: He is reading a paper [on SCIENCE]_F focus on a part of XP

- focus spreads over a sub-constituent, i.e. a prefix or a morphological part of a word which is not necessarily identical with the syllable carrying lexical stress, cf. (43):

- (43) Q: Is she Chinese?
 A: No, she is [JA]_Fpanese. (cf. Lambrecht 1994: 241)

While contrastive focus is located on the constituent viewed to bear the contrastive relation to another constituent, the determination of the domain of new-information focus reveals to be more problematic in our analysis. As was shown in the discussion on focus in section 3.1.3 as well as in the examples in this section, the identification of the focus domain of the utterance in current literature is based on (constructed) question-answer pairs. In striking contrast to that, no such methods are established for the identification of focus in sentences from natural running discourse as is the one in our analysis on Old High German data. However, we consider it possible to adopt the operational methods of focus detection established for dialogue sequences to texts from narratives, reports, etc. Let us assume that in line with the *quaestio*-theory by Klein and von Stutterheim (1992) each sentence of a running discourse is the adequate answer to an implicit question. In defining this question on the basis of the preceding discourse, one can identify those parts of the sentence that carry further the discourse and thus constitute the focus domain in sentences of running narrative texts.

4.2. Examples

The model proposed to capture the multitude of information-structurally relevant features of sentences constituents in data from the historic corpus of a language is demonstrated on the basis of sentences from the Old High German Tatian translation based on the edition of Masser (1994). For reasons described in section 2.1 above, the investigation primarily addresses sentences in which the Old High German text deviates from the sentence structure of the corresponding Latin original.

Below, some examples concerning the application of the model proposed as well as some comments to the assignment of features to the constituents are given.

Example 1

Lat. *Fuit in diebus herodis regis / iudeę quidam sacerdos / [...] / & uxor illi de filiabus aaron/& nomen eius elisab&h, / erant autem iusti ambo ante deum*

OHG *uuar [sic!] In tagun herodes thes cuninges / Iudeno sumer biscof / [...] / Inti quena Imo fon aarones tohterun / Inti ira namo uuas elisab&h / siu uuarun rehtiu beidu fora gote (T 25, 2)*

‘There was in the days of Herod, the king of Judea, a certain priest [...] and his wife of the daughters of Aaron, and her name was Elizabeth. And they were both righteous before God’

Sentence *siu uuarun rehtiu beidu fora gote*
they were righteous both before God

Lat.		<i>erant</i>	<i>autem</i>	<i>iusti</i>	<i>ambo</i>	<i>ante</i>	<i>deum</i>
OHG	<i>siu</i>	<i>uuarun</i>		<i>rehtiu</i>	<i>beidu</i>	<i>fora</i>	<i>gote</i>
informational status	giv						acc
TCS	top	comment					
aboutness	ab						
definiteness	def						def
syntactic realization	init						noninit
FBS	bgr	focus					
nif		nif					
cf							

Analysis:

In the OHG sentence, the subject pronoun *siu*-3pl ‘they’ is added against the Latin original and placed at the beginning of the clause, in front of the finite copula. The added constituent refers to entities already pre-established in the discourse. The application of the topic test *A says about X that X...* fits well to paraphrase the contents of the sentence and supports the interpretation of the constituent *siu* as the aboutness-topic of the sentence. All other features of topicality apply for *siu* as well, and we assign the category topic to it. The focus of the sentence covers the VP as the new information answering the implicit question “What about Zacharias and Elizabeth?”

Example 2

Lat. *Et pastores erant in regione eadem.* / *uigilantes & custodientes uigilias noctis / supra gregem suum*

OHG *uuarun thô hirta In thero lantskeffi.* / *uuahhante Inti bihaltante nahtuuahhta / ubar ero euuit*, (T 35, 29–31)

‘There were in the same country shepherds living out in the fields, keeping watch over their flock’

Sentence *uuarun thô hirta In thero lantskeffi*
 were then shepherds in that region

Lat.	<i>Et</i>	<i>pastores</i>	<i>erant</i>	<i>in</i>	<i>regione</i>	<i>eadem</i>
OHG	<i>uuarun</i>	<i>thô</i>	<i>hirta</i>	<i>in</i>	<i>thero</i>	<i>lantskeffi</i>
informational status			new			giv
TCS						
aboutness						
definiteness			indef			def
syntactic realization			noninit			nononit
FBS	focus					
nif	nif					
cf						

Analysis:

The sentence opens a new paragraph marked by concordance notes on the margins of the Latin part of the text as well as by capitalizing of the initial letter, cf. Masser (1994: 85). A new referent *hirta* ‘shepherds’ is introduced to the context; the expression denoting the new referent is placed in front of the finite copula in the Latin part and shifted after the finite verb in the Old High German sentence. Although the sentence contains the given and definite expression *in thero lantskeffi* ‘in the same country’, the general understanding of the sentence in this context does not support the intuition that it is meant not to provide further information on the region itself. The insertion of this referent into the topic-test *A says about X that X ...* does not yield a proper paraphrase of the sentence either. Following these considerations, we decide not to assign topic to *in thero lantskeffi*. Furthermore, the instance is a typical case of a presentational, all-focus sentence in which no topic-comment division applies. Thus, no topic category is assigned at all. As an introductory sentence of a new sub-episode, the entire sentence is focal.

Example 3

Lat. *Haec locutus sum uobis / ut In me pacem habeatis / In mundo presuram habebitis*

OHG *thisiu sprahih íu / thaz in mir habet sibba / In theru uueralti habet ir thrucnessi* (T 290, 7–9)

‘These things I have spoken to you, that in Me you may have peace. In the world, you will have turbulation’

Sentence *thaz in mir habet sibba*
that in me have-2Pl peace

Lat.	<i>ut</i>	<i>In</i>	<i>me</i>	<i>pacem</i>	<i>habeatis</i>
OHG	<i>thaz</i>	<i>in</i>	<i>mir</i>	<i>habet</i>	<i>sibba</i>
informational status			giv		new
TCS					
aboutness					
definiteness			def		def
syntactic realization					
FBS					
nif					
cf			cf_1		cf_2

Analysis:

In this adverbial clause the direct object *sibba* ‘peace’ is shifted to the post-verbal position in Old High German against the corresponding Latin structure. Two contrastive pairs are involved in this small discourse: one between *in me* and *In theru uueralti*, and another one between *sibba* and *thrucnessi*. For this reason, *in me* and *sibba* are analysed as two different instances of contrastive focus. The precise identification of the new information focus is difficult, though the most general implicit question fitting to the context is “Why have I spoken to you these words?” pointing at the entire subordinate clause as focal. Further subdivisions within this wide focus domain, e.g. taking the role of the finite verb as re-structuring device, are also possible but will be left out here.

5. Summary and further discussion

Pursuing the new perspectives which information structure reveals for the explanation of word order variation in early Germanic languages, we face the demand for a precise and systematic description of the information-structural value of sentence constituents and its impact on the syntactic

realization of these constituents in texts of the earliest Germanic records. Two main problems connected with this issue have been discussed: the possibility of isolating authentic material allowing for generalizations on true Germanic syntax as well as the methodology of applying criteria for information-structural description to data available in written form only. Concerning the first question in the light of investigating the role of information structure in OHG, it has been proposed to utilize the OHG Tatian translation as one of the earliest and largest data collections from this period and to restrict the analysis to a corpus of sentences deviating from the structure of the Latin original. Thus, the possible influence of the syntax of the Latin original on the structure of the Old High German translation is minimized.

Furthermore, we reflected the theoretical difficulties which information-structural research on historical data involves. We provided an overview of the controversial treatment of any information-structural category in current research and addressed the problems of assigning information-structural categories in written texts. It turned out that it is impossible to start from a clearly defined notion on the main categories in information-structural research and that contextual information and discourse organization are the most reliable sources for pragmatic features in historic data. According to these considerations we proposed a model based on the collection of prototypical features relevant for the information-structurally categories as shown on some selected instances of the Old High German Tatian translation.

The implementation of this model to a database and its statistical evaluation can help to come to grips with the enormous variation in the domain of word order and verb placement in early Germanic. It also allows for the detection of different combinations of features favoring special word-order patterns, and provides a way of looking for the distribution of patterns according to discourse-structural or other organizational principles. Furthermore, significant differences in the quantitative relations of competing patterns can be revealed, e.g. concerning different dialects, genres, or scribes. In this way, the statistical evaluation of a corpus enriched with the features of the proposed information-structural scheme is of enormous value for detecting some ordering principles in an apparently unordered system and for localizing domains in which the establishment of general rules first applied.

Acknowledgments

The present paper is a revised version of a talk given in the Workshop ‘Information Structure and Language Change’ in Berlin (Germany), September 30th–October 1st, 2005 organized by Project B4 of Collaborative Research Center 632 ‘Information Structure: the linguistic means for structuring utterances, sentences and texts’ of Potsdam University and Humboldt University Berlin. We thank all participants of the workshop for questions and discussions, and an anonymous reviewer for comments on an earlier version of the paper.

Notes

1. We refer to the work of project B4 (principal investigators Karin Donhauser and Roland Hinterhölzl) as part of Collaborative Research Center 632 “Information Structure” at Humboldt University Berlin launched in July 2003 and funded by the German Research Foundation (DFG). For a more detailed description on the issues of research see <http://www.linguistik.hu-berlin.designato.de/sprachgeschichte/forschung/informationsstruktur/index.php> or <http://www.ling.uni-potsdam.de/sfb/>.
2. Here we find on four pages alone eleven violations of the translation principle described above. A number of seven instances, as mentioned by Dittmer and Dittmer (1998: 24–25), who did not examine the whole text, is proven to be too low, cf. Fleischer et al. (2008: 220), who account for 46 instances in total.
3. Sonderegger (1965) provides a similar attempt to classify the different kinds of syntactic divergences in Old High German translations.
4. For the distribution of the Old High German subject pronoun cf. Eggenberger (1961).
5. As Masser (1991) seems to suppose for scribe ϵ .
6. The example is taken from Chafe (1976: 30).
7. Cf. the homepage of Project 9 at the Centre for General Linguistics, Typology and Universals Research (ZAS) Berlin <http://www.zas.gwz-berlin.de> [20.07.2006].

References

Bästlein, Ulf Christian

- 1991 *Gliederungsinitialen in frühmittelalterlichen Epenhandschriften. Studie zur Problematik ihres Auftretens, ihrer Entwicklung und Funktion in lateinischen und volkssprachlichen Texten der Karolinger- und Ottonenzeit.* Europäische Hochschulschriften. Reihe I. Deutsche Sprache und Literatur; 1167. Frankfurt a. M.: Peter Lang.

Bean, Marian C.

- 1983 *The Development of Word Order Patterns in Old English.* Croom Helm Linguistic Series. London: Croom Helm.

Behaghel, Otto

- 1932 *Deutsche Syntax. Eine geschichtliche Darstellung. Band IV. Wortstellung, Periodenbau.* Heidelberg: Carl Winter.

Chafe, Wallace L.

- 1976 Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In *Subject and Topic*, ed. Charles N. Li, 25–55. New York: Academic Press.

Clark, Herb H. and Haviland, Susan E.

- 1977 Comprehension and the given-new contrast. In *Discourse Production and Comprehension*, ed. Roy O. Freedle, 1–40. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Denschewa, Emilia

- 1987 Zur sprachlichen Eigenständigkeit der althochdeutschen Tatian-Übersetzung in bezug auf den Gebrauch des Infinitivs. *Beiträge zur Erforschung der deutschen Sprache* 7: 207–232.

Dik, Simon C.

- 1989 *The Theory of Functional Grammar. Part I: The Structure of the Clause.* Dordrecht: Foris Publications.

Disse, Andreas

- 1998 *Informationsstruktur im biblischen Hebräisch. Sprachwissenschaftliche Grundlagen und exegetische Konsequenzen einer Korpusuntersuchung zu den Büchern Deuteronomium, Richter und 2 Könige.* Arbeiten zu Text und Sprache im Alten Testament; 56/Teil 1. St. Ottilien: EOS Verlag.

Dittmer, Arne and Dittmer, Ernst

- 1998 *Studien zur Wortstellung – Satzgliedstellung in der althochdeutschen Tatianübersetzung.* Göttingen: Vandenhoeck & Ruprecht.

Eggenberger, Jakob

- 1961 *Das Subjektspronomen im Althochdeutschen.* Chur: Sulser.

- Firbas, Jan
 1966 On defining the 'theme' in functional sentence analysis. In *Travaux Linguistiques de Prague*, vol. I, 267–280. University of Alabama Press.
- Fleischer, Jürg
 2006 Zur Methodologie althochdeutscher Syntaxforschung. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 128: 25–69.
 this vol. Paleographic clues to Old High German prosody? – Accents, word separation, and related phenomena in Old High German manuscripts In *New Approaches to Word Order Variation and Change in the Germanic Languages*, eds. Roland Hinterhölzl and Svetlana Petrova. Berlin: Mouton de Gruyter.
- Fleischer, Jürg, Hinterhölzl, Roland and Solf, Michael
 2008 Zum Quellenwert des althochdeutschen Tatian für die Syntaxforschung. *Zeitschrift für germanistische Linguistik* 36: 210–239.
- Floor, Sebastian Jonathan
 2004 From Information Structure, Topic and Focus, to Theme in Biblical Hebrew Narrative. Ph.D. University of Stellenbosch.
- Frey, Werner
 2005 Notes on the syntax and the pragmatics of German Left Dislocation. In *The Syntax and the Semantics of the Left Sentence Periphery*, eds. Horst Lohnstein and Susanne Trissler, 203–233. Berlin: Mouton de Gruyter.
- Gundel, Jeanette K.
 1988 Universals of topic-comment structure. In *Studies in Syntactic Typology*, eds. Michael Hammond, Edith A. Moravcsik and Jessica R. Wirth. Amsterdam: John Benjamins.
- Halliday, Michael A. K.
 1967 Notes on Transitivity and Theme in English, II. *Journal of Linguistics* 3: 199–244.
- Heimerdinger, Jean-Marc
 1999 *Topic, Focus and Foreground in Ancient Hebrew Narratives*. Sheffield: Sheffield Academic Press.
- Hinterhölzl, Roland
 2004 Language Change versus Grammar Change: What diachronic data reveal about the distinction between core grammar and periphery. In *Diachronic Clues to Synchronic Grammar*, eds. Eric Fuß and Carola Trips, 131–160. Amsterdam and Philadelphia: John Benjamins.
 this vol. The role of Information Structure in word Order Variation and word Order Change. In *New Approaches to Word Order Variation and Change in the Germanic Languages*, eds. Roland Hinterhölzl and Svetlana Petrova. Berlin: Mouton de Gruyter.

- Hockett, Charles F.
 1958 *A Course in Modern Linguistics*. New York: Macmillan.
- Hopper, Paul J.
 1979a Some Observations on the Typology of Focus and Aspect in Narrative Language. *Studies in Language* 3.1: 37–64.
 1979b Aspect and Foregrounding in Discourse. In *Syntax and Semantics*, ed. Talmy Givón, 213–241. San Diego et al.: Academic Press, INC.
- Ingenbleek, Theodor.
 1880 Über den Einfluss des Reimes auf die Sprache Otfrids besonders in Bezug auf die Laut- und Formenlehre. Mit einem Reimlexikon zu Otfrid. Straßburg: Karl J. Trübner.
- Jacobs, Joachim
 2001 The dimensions of topic-comment. *Linguistics. An interdisciplinary journal of the language sciences* 39/4: 641–681.
- Kleiber, Wolfgang
 2004 Otfrid von Weißenburg. Evangelienbuch. Band I: Edition nach dem Wiener Codex 2687. Teil 2: Einleitung und Apparat. Tübingen: Niemeyer
- Klein, Wolfgang, and Stutterheim, Christiane von
 1992 Textstruktur und referentielle Bewegung. *Zeitschrift für Literaturwissenschaft und Linguistik* 86: 67–92.
- Krifka, Manfred
 2007 Basic notions on Information Structure. In *The notions of Information Structure*, eds. Caroline Féry et al., 13–56. Potsdam: Universitätsverlag.
- Kuno, Susumo
 1972 Functional Sentence Perspective: a Case Study from Japanese and English. *Linguistic Inquiry* 3: 269–320.
 1978 Generative Discourse Analysis in America. In *Current Trends in Textlinguistics*, ed. Wolfgang U. Dressler, 275–294. Berlin and New York: de Gruyter.
- Lambrecht, Knud
 1994 *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Lühr, Rosemarie
 2005 Der Einfluß der klassischen Sprachen auf die germanische Grammatik. In *Sprachkontakt und Sprachwandel. Akten der XI. Fachtagung der Indogermanischen Gesellschaft*, eds. Gerhard Meiser and Olav Hackstein, 341–362. Wiesbaden: Reichert.

Masser, Achim

- 1991 Die lateinisch-althochdeutsche Tatianbilingue des Cod. Sang. 56. *Nachrichten der Akademie der Wissenschaften in Göttingen, I. Philologisch-historische Klasse*, 1991, 3. Göttingen: Vandenhoeck & Ruprecht.
- 1997a Syntaxprobleme im althochdeutschen Tatian. In *Semantik der syntaktischen Beziehungen. Akten des Pariser Kolloquiums zur Erforschung des Althochdeutschen 1994*, ed. Yvon Desportes, 123–140. Heidelberg: Carl Winter.
- 1997b Wege zu gesprochenem Althochdeutsch. In *Grammatica Ianua Artium. Festschrift für Rolf Bergmann zum 60. Geburtstag*, eds. Elvira Glaser and Michael Schläefer, 49–70. Heidelberg: Carl Winter.

Masser, Achim ed.

- 1994 Die lateinisch-althochdeutsche Tatianbilingue Stiftsbibliothek St. Gallen Cod. 56. Göttingen: Vandenhoeck & Ruprecht.

Molnár, Valéria

- 1993 Zur Pragmatik und Grammatik des TOPIK-Begriffs. In *Wortstellung und Informationsstruktur*, ed. Marga Reis, 155–202. Tübingen: Max Niemeyer.

Musan, Renate

- 2002 Informationsstrukturelle Dimensionen im Deutschen: Zur Variation der Wortstellung im Mittelfeld. *Zeitschrift für Germanistische Linguistik* 30: 198–221.

Prince, Ellen F.

- 1981 Toward a Taxonomy of Given-New Information. In *Radical Pragmatics*, ed. Peter Cole, 223–255. New York: Academic Press.

Reinhart, Tanya

- 1981 Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27: 53–94.

Ries, John

- 1880 *Die Stellung von Subject und Prädicatsverbum im Heliand: Nebst einem Anhang metrischer Excuse*. Straßburg/London: Karl J. Trübner.

Rooth, Mats

- 1985 Association with Focus. Ph.D. University of Massachusetts at Amherst.

Ruhfus, Wilhelm

- 1897 *Die Stellung des Verbuns im althochdeutschen Tatian*. Dortmund: Ruhfus.

Sasse, Hans-Jürgen

- 1995 “Theticity” and VS order: a case study. *Sprachtypologie und Universalienforschung* 48, 3–31.

- Sievers, Eduard
1960 Reprint. *Tatian*. 2. neubearb. Aufl. Paderborn: Schöningh. Original edition, Paderborn: Schöningh, 1892.
- Simmler, Franz
1998 Makrostrukturen in der lateinisch-althochdeutschen Tatianbilingue. In *Deutsche Grammatik: Thema in Variationen*, eds. Karin Donhauser and Ludwig M. Eichinger, 299–335. Heidelberg: Carl Winter.
- Sonderegger, Stefan
1965 Frühe Übersetzungsschichten im Althochdeutschen. In *Philologia Deutsch*, eds. Werner Kohlschmitt and Paul Zinsli, 101–114. Bern: Francke.
- Stockwell, Robert P.
1984 On the history of the verb-second rule in English. In *Historical Syntax*, ed. Jacek Fisiak, 575–592. Berlin / New York / Amsterdam: Mouton.
- Szendrői, Kriszta
2004 Focus and the interaction syntax and pragmatics. *Lingua* 114: 229–254.
- Weil, Henri
1879 L'ordre des mots dans les langues anciennes comparées aux langues modernes. *Question de grammaire générale*. 3rd edition. Paris: F. Vieweg. [1844]
- Zybatow, Gerhild
1999 Informationsstruktur im Russischen. In *Sprachspezifische Aspekte der Informationsverteilung*, ed. Monika Doherty, 61–89. Berlin: Akademie Verlag.