

# Phrasing it differently

Kim Gerdes & Sylvain Kahane

Lattice, Université Paris 7, case 7003, 75251 Paris Cedex 5

kim@linguist.jussieu.fr sk@ccr.jussieu.fr

*Si tu prends un arbre, un vrai arbre, pas un arbre de constituants, qui n'existe pas, ...*<sup>1</sup>

(Igor Mel'cuk, conference at the ENS, Paris, May 28 2002)

**Abstract.** This paper investigates the notion of phrase in non phrase structure grammars. Following Tesnière and Mel'cuk, we defend the idea that the word order must be separated from the syntactic representation proper and that phrases only intervene when word order is at play. We try to characterize a new notion we call *topological phrase* (partially inherited from the classical topological model for German) and distinguish it from the classical notion of phrase in X-bar Syntax. Our discussion is illustrated by the puzzling case of German word order for which we propose a simple and powerful grammar giving us all the possible word orders and topological phrase structures of verbal syntax. This discussion of the notion of *phrase* opens a new perspective for the comparison of the entire architectures of Chomskyan and Mel'cukian linguistic models.

## 1 Introduction

We have written this article for the 70<sup>th</sup> birthday of Igor Mel'cuk. The ideas that we present here owe a lot to his work and to the discussions we had with him. Our goal in this article is to develop a topic, which he has frequently put forward in our presence, but which, we believe, has not been sufficiently reflected in his writings.

Igor Mel'cuk may presently be the most emblematic figure of the dependency approaches to linguistics and dependency has often been opposed to constituency (Hudson 1980, Mel'cuk 1988). Just as others in the dependency community, Igor Mel'cuk rejects the notion of *syntactic* constituency and its prominent role in the description of language. Following Tesnière 1959, he argues that the syntactic representation of a sentence should not include word order. Rather than rejecting the notion of linguistic con-

---

<sup>1</sup> “Take a tree, a real tree, not a constituent tree that does not exist, ...”

stituency completely, he thinks that constituency is coupled with word order and should appear in a representation separated from the syntactic structure, that he calls the (*deep*) *morphological representation*. This representation “specifies the form of a particular sentence in terms of the word forms and the phrases that constitute it and their linear order, without regard to the internal organization of the word forms themselves.” (Mel’cuk 1988: 69)

Yet, the notions of “phrase” or “morphological constituency” are the stepchildren of the Meaning-Text Theory. They appear sporadically in Igor Mel’cuk’s work, but have not been developed in detail. His most advanced work in this direction is probably Mel’cuk 1965 on word order in Russian. In Mel’cuk 1988, the morphological constituents appear only in one example (Figure 2-7 page 71) where he gives the deep morphological representation of a Russian sentence. The deep morphological representation consists of two structures:

- an ordered string of deep morphological representations of all the word forms that make up the sentence, given in the linear order they actually have in the sentence;
- a cutting up of the string with a prosodic markup composed of “pauses, intonation contours, and the like” (*ibid.*: 71).

We find it necessary to modify form and function of the (deep) morphological representation. This paper proposes to replace this morphological representation by a structure that we call *topological representation*, including an ordered hierarchy of constituents, which we will contrast with the constituents of usual phrase structure based grammars.

In Section 2 we establish the foundations of our approach. Section 3 applies our ideas to German and proposes a grammar for its syntax-topology interface. Section 4 compares different topological phrase structures that can correspond to a given dependency tree. Section 5 sharpens up our German grammar in order to show how our approach works out smoothly even subtle details of German word order. In Section 6 we tackle the question of the characterization of topological phrases and we show their prosodic existence. Section 7 proposes a general comparison of the architecture of Mel’cukian and Chomskyan models of language and the status of phrases.

## 2 Problematics

We distinguish two types of linguistic constituents: Syntactic constituents and topological constituents.

- *Syntactic constituents* are (maximal) syntactic projections of lexical heads. They are directly related to the notion of syntactic dependency. A syntactic constituent corresponds to the ordered string of the nodes dominated by a node in the dependency tree. Conversely, a syntactic dependency expresses the relation between the lexical head of a constituent and the lexical head of a subconstituent, leaving aside the linear order of these two elements.

- *Topological constituents* are linearly ordered groups of words appearing naturally in the layout of the word order of the sentence. They are the bags of words considered in the computation of word order, motivated both by the syntactic dependency and the communicative groups.<sup>2</sup> They are necessarily continuous segments of the sentence and are generally prosodically marked. We think that only these constituents should be legitimately called *phrase* in the Saussurean sense of the word (fr. *syntagme*).

This paper defends the following ideas:

- The usual syntactic phrase structure description attempts to encode two kinds of information: syntactic dependency and topological constituency.
- Syntactic dependency and topological constituency are independent (although related) notions. The syntactic dependency expresses the syntactic hierarchy of words, and not their linear order. The topological phrase structure expresses the order and the grouping of the words.
- A description of language is a description of the relation between meanings and forms of the language. Syntactic dependency and topological constituency are intermediate structures between meanings and forms.
- The syntactic dependency structure is an intermediate description closer to the meaning than the topological phrase structure, which is closer to the form of the language.
- On the way from the meaning to the form of a sentence, we pass through the description of the sentence as a syntactic dependency structure before passing through the description of the sentence as a topological phrase structure.
- The relation between the syntactic dependency structure and the topological phrase structure is direct. Its description is an important part of the syntax of the language. It is an independent module of our language model.
- The relation between the syntactic dependency structure and the topological phrase structure depends on communicative information: The constituent structure is constrained both by the syntactic dependency structure and the communicative structure.

---

<sup>2</sup> The communicative structure is more commonly known under the term *information structure* (Lambrecht 1994). Its main goal is to distinguish what the proposition is about (theme, presupposition, topic, ...) from the new information on this subject (rheme, focus, comment, ...). Communicative groups are entities carrying a communicative markup that can appear on different levels of linguistic representation. Refer to Mel'cuk 2001 for details.

- The syntactic dependency structure and the communicative structure have to be considered at the same intermediate level of the language model and are thus both part of the syntactic representation.
- The constituents carry directly the communicative markup. At this level of representation, the communicative information is no longer separated, but merged into a unified structure, the constituent structure, communicative information being a markup on the constituents.
- The topological constituents are in direct correspondence with the units of the prosodic structure, created in the following step of the language production process.

Most of these ideas correspond to the postulates underlying the Meaning Text Theory. Our restrictive use of phrase structures fits them into the Meaning Text framework and corroborates these postulates. We find it necessary to propose two modifications on the so-called morphological level.

First, the intonation contours should not be decided on before the word has been fully developed. What they should carry is only discrete information of communicative and topological nature.

Second, the morphological structure is not a flat sequence of constituents. When linearizing, we obtain naturally a hierarchy of constituents rather than just a sequence. We should maintain this information since this hierarchy of constituents corresponds on the next level of representation to a hierarchy of prosodic contours, which will be superposed to obtain the final intonation.

We consider that the terms of *morphological representation* and *morphological phrase/constituent* lead to misinterpretation because *morphological* usually evokes a word-internal property. We prefer to use the term *topological representation* and *topological phrase/constituent*.<sup>3</sup>

### 3 Topological Phrase Structure for German

German appears as an interesting choice for our study for two reasons. First, German surface phenomena are complex and widely studied, and different formal analyses of the German word order have been proposed outside of the common Chomskyan framework (Kathol 1995, 2000, Debusmann and Duchier 2001, Gerdes and Kahane 2001). Second, although German is considered as a “free word order language”, its word order obeys strict constraints. In English or French, the syntactic and the topological constitu-

---

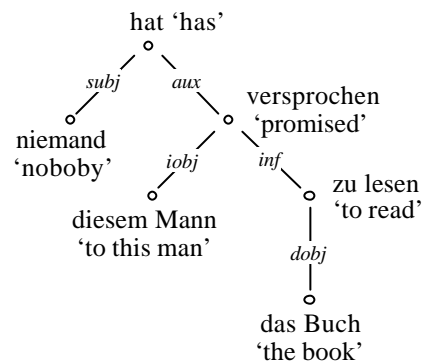
<sup>3</sup> In our topological representations, words must be represented by their morphological representation, that is their decomposition into morphemes. Nevertheless, in the following, we do not consider the morphology and we use words in topological representations, as well as in syntactic representations.

ents often coincide, and it is difficult to discern a distinctive topological structure. In less configurational languages like Russian, the surface order is nearly exclusively determined by the communicative structure. German appears as one of the most interesting cases because surface order depends strongly on both the syntactic structure (e.g. finite verbs go in verb second or verb final position, depending on whether they are governed) and the communicative structure (e.g. the content of the *Vorfeld*).<sup>4</sup>

This section begins with a presentation of different possibilities of word order in German (Section 3.1). It is followed by a presentation of the classical topological model and the particularities of our approach (Section 3.2). We finish with the basic rules of our German grammar (Section 3.3). The grammar will be completed in Section 5.

### 3.1 *Different order phenomena in German*

Let us consider an example to show the different possibilities of German word order. The syntactic dependency tree of Fig. 1, which will be our reference example, has a few dozens linearizations, among them the sentences in (1).<sup>5</sup>



**Figure 1: Dependency tree of the sentences in (1)**

---

<sup>4</sup> We call *Vorfeld* the unique position before the finite verb of a declarative sentence, making German a verb second (V2) language. The *Vorfeld* can accommodate a theme as well as a rheme (see Section 6).

<sup>5</sup> We do not give the internal structure of the noun phrase. The syntactic dependency between the noun and the determiner goes off the topic and is certainly too controversial to be discussed on the fly (see Note 21).

- (1) a. *Niemand hat diesem Mann das Buch zu lesen versprochen*  
 b. *Diesem Mann hat das Buch niemand zu lesen versprochen*  
 c. *Das Buch zu lesen hat diesem Mann niemand versprochen*  
 d. *Diesem Mann hat niemand versprochen, das Buch zu lesen*  
 e. *Diesem Mann hat, das Buch zu lesen, niemand versprochen*  
 f. *Zu lesen hat das Buch diesem Mann niemand versprochen*  
 g. *Das Buch hat niemand diesem Mann versprochen zu lesen*  
 h. *Das Buch hat zu lesen niemand diesem Mann versprochen*
- ‘Nobody promised this man to read the book.’

Following Tesnière 1959 and Mel’cuk 1988, we consider the finite verb, auxiliary or full verb, as the syntactic head of the sentence, depicted as the root node of the dependency tree. Only finite verbs are considered to govern a subject as they agree with it.

Our goal is to describe all possible word orders for a well-formed syntactic dependency tree. Note that the well-formedness of syntactic dependency trees (for example the completeness of the subcategorization frames) is taken care of by the semantics-syntax interface and is not of our interest in this study.

The finite verb *hat* takes the second position in all the sentences corresponding to the dependency tree of Figure 1. This means that there is exactly one constituent before the main verb of the sentence. In the most common cases as (1a,b), the non-finite verbal dependent of the main verb is in sentence-final position and forms a *verb cluster* with its own non-finite verbal dependent. The latter non-finite verbal dependent can again be joined by another non-finite verbal dependent in a way that the verb cluster is made up of a subordination chain of non-finite verbs (called *hypotactic chain* by Bech, 1955:26).<sup>6</sup> In the verb cluster, the governor ordinarily follows its dependent, but the reverse order is possible in some constrained cases, which we will study below. In our reference example, the final verb cluster is *zu lesen versprochen* ‘to read promised’. The order of the three nominal groups, although they depend on different verbs, is syntactically unconstrained: Any of them can be in front of the finite verb *hat* and the two remaining can appear in any order between *hat* and the verb cluster. The constituent in front of the finite verb *hat* can be of a more complicated nature: It can be the infinitive clause *das Buch zu lesen* (the projection of *zu lesen*) as in (1c). This phenomenon is called *VP-fronting*. The same constituent can also appear behind the “final” verb *versprochen* as in (1d). This is called *extraposition*. We speak of *intraposition* when the

---

<sup>6</sup> In a subordinate clause, the main verb is in the clause-final position and usually forms a verb cluster with its non-finite verbal dependent (*Ich glaube, dass niemand diesem Mann das Buch zu lesen versprochen hat* ‘I think that nobody promised this man to read the book’).

same constituent appears between the finite verb and the “final” verb *versprochen* as in (1e). Another possibility is that *zu lesen* takes the position before the finite verb *hat* without its dependent *das Buch* as in (1f). This case is usually referred to as *partial VP-fronting*. Here, the order among *das Buch* and the other nominal complements is free: If *das Buch* does not join the infinitive clause, it behaves just as the dependents of the higher verbs. We call this *emancipation*. Emancipation is also possible when the infinitive is not fronted but extraposed (1g) or intraposed (1h).

Note that *das Buch zu lesen* in (1a) could also be analyzed as an intraposed infinitive clause as in (1e). Most phrase structure based approaches consider this “verb-phrase-embedding” analysis as basic, and (1b), where this embedding analysis is not possible, as deviant from the basic structure denoted by terms like *scrambling*. In Section 6 we will defend that the sentence (1a) has these two analyses (and even a third) that are prosodically marked and communicatively motivated. In our analysis, these orders are independently generated from the syntactic dependency tree on communicative grounds.

### 3.2 *Topological model*

Our approach is based on the classical *topological model* first introduced in the description of German (Drach 1937, Bech 1955). Such an approach has also been proposed for the description of word order in non Germanic language such as Ancient French (Skårup 1975) and Warlpiri (Donohue and Sag 1999). The topological model has been satisfactorily implemented in HPSG (Kathol 1995, 2000) and in dependency grammars (Debusmann and Duchier 2001, Gerdes and Kahane 2001). The initial idea of the topological model is to consider that a sentence is a template-like sequence of different fields each being able to host different types of constituents. In Gerdes and Kahane 2001, we have extended this approach by explicitly considering a topological phrase structure and applying recursively classical ideas from the first works on the topological model: In our approach, placing an element in linear order means creating topological constituents; each topological constituent is internally organized as a sequence of fields which can in turn host topological constituents and so forth.

Minimal constituents are lexical, named after the part of speech of the word. Around its minimal constituent, the word can then create larger topological constituents, capable of hosting some of the word’s dependents. The highest constituents words can (but do not have to) create are called *domains*. Non-lexical constituents possess a sequence of fields: For example, the *main domain* (opened by the finite main verb) is the underlying pattern of the German declarative sentence, and it consists of the following sequence of five fields: [*Vorfeld*, *left bracket*, *Mittelfeld*, *right bracket*, *Nachfeld*]. A domain resembles a box whose ordered compartments, called fields, can themselves accommodate new boxes.

We then have three types of rules:

1. *Constituent creation rules* give the types of constituents a word can create and specifies which field of the constituent it occupies;
2. *Constituent description rules* describe the ordered list of fields the constituent consists of and indicate whether a field can or must accommodate one or more constituents;
3. *Constituent placement rules* indicate into which field a word can go – depending on the position and the constituent of its governor.

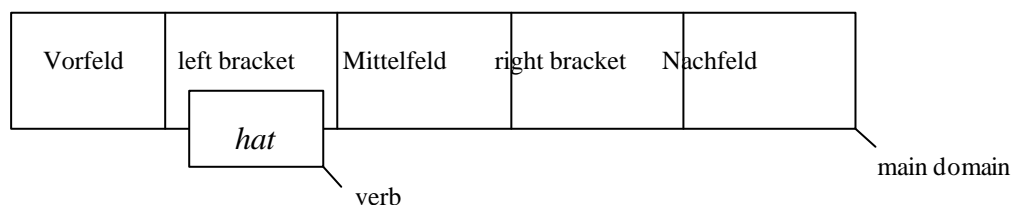
These rules constitute the principles of our syntactic module, which realizes the syntax-topology interface. We postulate that the syntactic module of any language could be described in these terms. For English, the topological structure could be very similar to the traditional syntactic phrase structure, including a verbal phrase with the verb and its object.<sup>7</sup> We will see that German shows serious mismatches between the syntactic constituents and the topological phrases.

In the following we will describe the rules for German declarative sentences leaving aside the noun-internal structure.

### 3.3 *Word order rules for German*

We have established the following rules for the linear order of verbs and their dependents:

- Every node of the syntactic dependency tree creates a lexical constituent named after its part of speech.
- The main finite verb creates a *main domain*, consisting of the following sequence of five fields: [*Vorfeld*, *left bracket*, *Mittelfeld*, *right bracket*, *Nachfeld*]. It takes the second position of this domain, the left bracket. A verb in this position is also called V2.

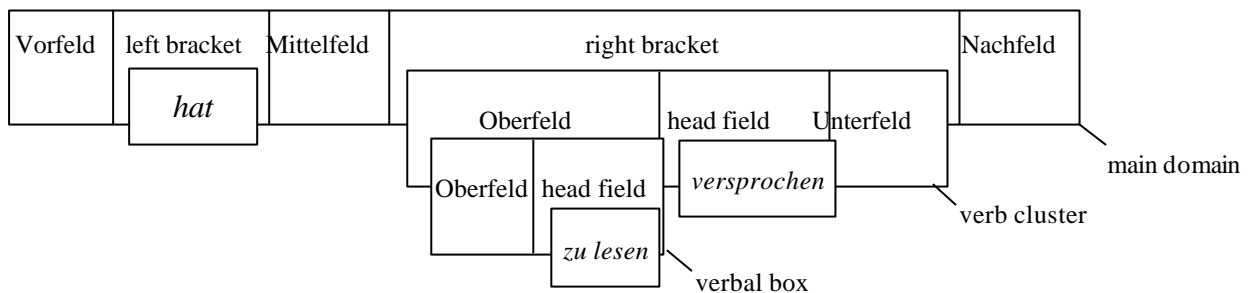



---

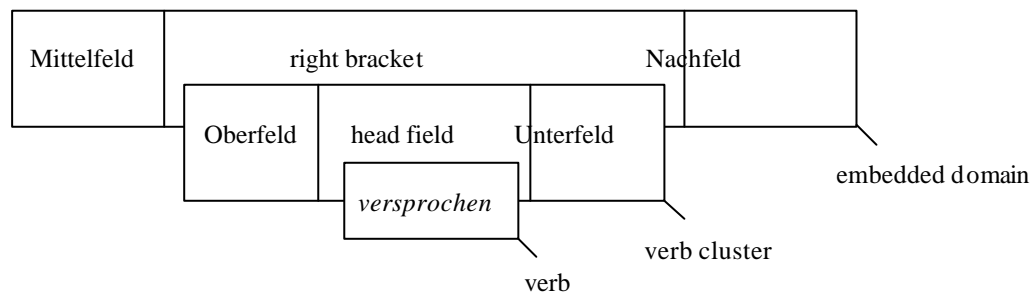
<sup>7</sup> Another debate, which is not of our concern here, regards the legitimacy of the syntactic verbal phrase in the syntactic representation for English and non-English languages (see for instance Abeillé 1996-1997 for the French). We think that this question is largely internal to the Chomskyan framework and has no theoretical support in dependency grammars. Moreover, we suspect the “VP-hypothesis” to be partially instigated by the predominant role of English in linguistic studies.



- To obtain a grammatical declarative sentence, the Vorfeld and the left bracket of the final topological structure must contain exactly one constituent. The right bracket can host at most one constituent. The number of constituents that the Mittelfeld and the Nachfeld take is not constrained.
- Any dependent verb opens a constituent called the *verb cluster*. This constituent has three fields, the *Oberfeld*, the place for the lexical constituent, and the *Unterfeld*. The Oberfeld and, under some conditions, the Unterfeld will be a possible place for a non-finite verbal dependent (see Section 5.1).<sup>8</sup>



- The verb cluster goes into the right bracket of a main domain. Two choices are possible: the verb cluster opened by the verbal dependent of the main verb goes into the right bracket of the main domain (opened by the main verb) or the dependent verb opens itself a domain and the verb cluster goes into the right bracket of this subdomain.
- The domain opened by a non-finite verb, called an *embedded domain*, consists of three fields: *Mittelfeld*, *right bracket* and *Nachfeld*. The verb cluster takes the right bracket.



<sup>8</sup> To avoid confusion, note that our usage of the terms Oberfeld and Unterfeld does not correspond to their original meaning from Bech 1955: Bech's Oberfeld was only taken by the auxiliaries in the case of Oberfeldumstellung (see Section 5.1 for details); the verbs in common order are limited to the Unterfeld. We only keep the order of Oberfeld before Unterfeld and add a head field between the two.

- All dependents of the main finite verb can create a subdomain that must go in one of the three *major fields* (Vorfeld, Mittelfeld, Nachfeld). The particularity of the verbal dependent is to have the choice not to open a subdomain. The choice of the major field the subdomain can occupy is not free, but obeys syntactic and communicative restrictions.
  - The domain created by the past participle and the bare infinitive is the most constrained: It has to go into the Vorfeld of the main domain.
  - The domain created by a *zu*-infinitive obeys no restrictions: It can go in any major field.<sup>9</sup>
  - The domain of sentential complements can go in either the Vorfeld or the Nachfeld.
  - Nominal and prepositional domains can go in any of the major fields, including the Nachfeld under heaviness constraints.
- *Particles* (traditionally called *separable verbal prefixes*), such as the *an* of *anfangen* ‘begin’), behave exactly like verbs and can go into the right bracket (when the governor directly follows its particle, it is customary to write the particle and the governor as one word).<sup>10</sup> The particle opens a position for a verbal dependent as in (2):
 

(2) *Er fängt gleich zu schreien an*  
       he begins right-away to shout AN(prefix of begin)  
       ‘He begins to shout right away’
- Some non-verbal dependents, such as predicative adjectives and nouns governed by a copular or support verb, can go into the right bracket. In contrast to verbs and particles, these elements do not usually open up a new position for their dependents, which consequently have to be placed somewhere else.

---

<sup>9</sup> As is common practice, we treat the *zu* of the *zu*-infinitives (and the corresponding *te* in Dutch) as an inflectional marker and not as a separate word, because, contrarily to English, no lexical element can ever be inserted between it and the verb and, in particle verb constructions, it is customary in writing to attach them to the verb.

<sup>10</sup> The capacity to form a new domain depends on the semantics of the particle: Just like bare infinitives that can be fronted only into the Vorfeld and only in order to express a contrast, particles can only open a new domain in the Vorfeld and only if they can communicatively be opposed to another particle.

- (i) *Auf hat er die Tür nicht gemacht, aber zu.*  
       open has he the door not done, but close  
       ‘He did not OPEN the door, he CLOSED it.’

- A subdomain can go in a major field of any domain containing its governor. We call *emancipation* a case when it does not go in the first domain containing its governor. In (1f-h), *das Buch* ‘the book’ is emancipated from the domain opened by its governor *zu lesen* ‘to read’ and is in a major field of the main domain opened by the auxiliary *hat*.

The above fragment of German grammar is presented in a formalized way in Gerdes & Kahane 2001. In Section 5, we will extend the grammar to the verb cluster’s internal structure and to sentential complements and relative clause.

## 4 Comparison of different topological phrase structures

In this section, we will compare different groupings that can appear in the topological phrase structures associated with a given syntactic dependency tree. Our reference example remains the tree of Figure 1. For this dependency tree, we have automatically calculated 144 topological phrase structures (we obtain 188 structures if we include sentences with the expletive *es* and 666 structures if we allow nominal extraposition into the *Nachfeld*).<sup>11</sup> These numerous topological phrase structures can be classified by the domains appearing in them. Independently of the linear position that the constituents finally take, we can distinguish cases without embedded domains, cases where some of the verbs create embedded domains, and cases where these embedded domains do not contain all of the dependents of the domain-governing verb. We start with the easiest cases, the flat structures.

### 4.1 Flat topological phrase structures

The least marked topological phrase structures of a German sentence do not have any subdomains. In this case, for our reference example, the subordinated verbs *versprochen* ‘promised’ and *zu lesen* ‘to read’ will go into the right bracket of the main domain headed by *hat* ‘has’. As no new domain has been created, all the non-verbal dependents of the three verbs share the major fields and are ordered independently of the verbal subordination hierarchy, obeying only communicative constraints.

Abstracting away from the linear order, we can represent the domain structure projected onto the dependency tree as in Figure 2: We superimposed the dependency tree to the topological groups. The constituents that occupy the left and right brackets are represented by shaded ovals.

---

<sup>11</sup> This has been realized with the program DepLin developed by Kim Gerdes ([tana.linguist.jussieu.fr/~kim/deplin](http://tana.linguist.jussieu.fr/~kim/deplin)), which uses the German grammar presented here (and formalized in Gerdes & Kahane 2001) and associates to a given syntactic dependency tree all its possible topological phrase structures.

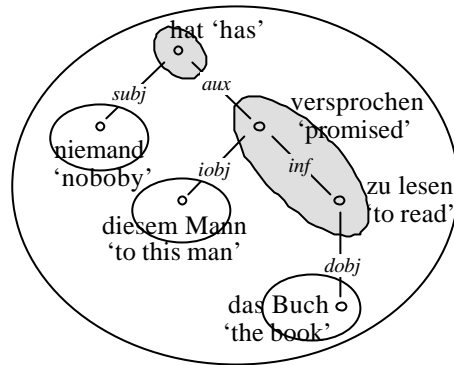
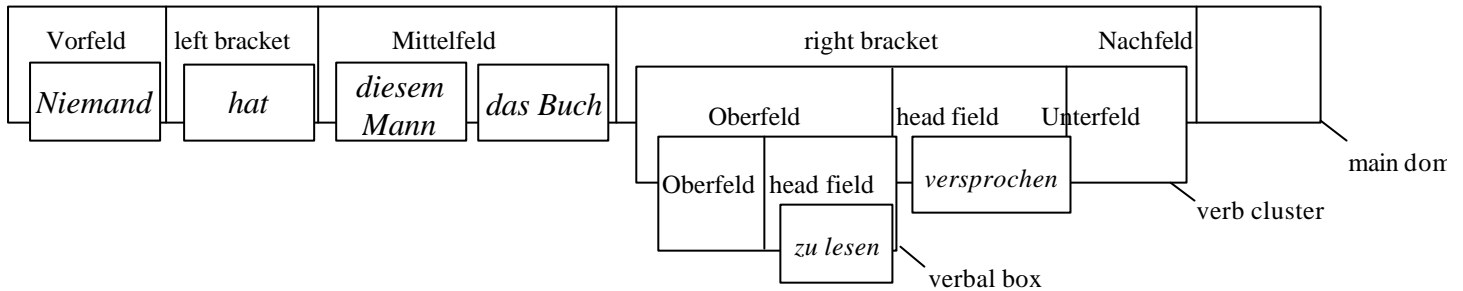


Figure 2: A phrase structure without embedded domains corresponding to (1a,b)

In the flat grouping, the three non-verbal phrases, *niemand* ‘nobody’, *diesem Mann* ‘to this man’, and *das Buch* ‘the book’, are on the same domain level; one of them has to take the Vorfeld, the other two will go into the Mittelfeld. We thus obtain 6 possible



orders, among them (1a) and (1b). The topological structure of (1a) is shown here:

There are nevertheless some general restrictions on the relative constituent order in the Mittelfeld. We do not consider these rules here (see for instance Lennerz 1977, Uszko-reit 1987), but we want to insist on the fact that the order of the constituents depends very little on their hierarchical position in the syntactic structure. The linear order is not free but depends on restrictions that weigh more heavily than the hierarchical position: pronominalization, focus, heaviness, etc. Dutch has the same basic topological structure as German, but the order in the Mittelfeld is very constrained syntactically. It is certainly linked to the fact that Dutch has lost grammatical case (except on pronouns). For a simplified description of the order in the Dutch Mittelfeld, we have to attach to each complement placed in the Mittelfeld its height in the syntactic dependency tree, and linearize them in descending order.

The fact that a verbal projection (i.e. the verb and all of its direct and indirect dependents) does not in general form a continuous phrase, unlike in English and French, is called *scrambling* (Ross 1967). This terminology is based on a conception of syntax we reject, which supposes that word order is primarily guided by the syntactic hierarchy (i.e. every projection of a given element forms a phrase) and that any deviation from this constitutes a problem. It presupposes that there is “standard order” totally reflecting the

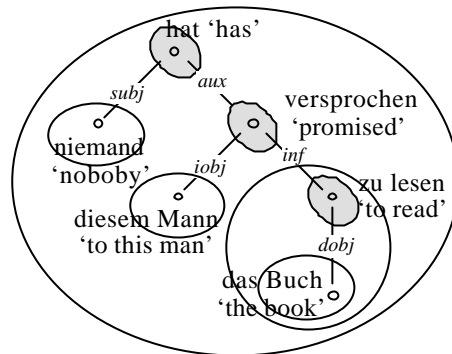


Figure 3: A phrase structure with an embedded domain corresponding to (1a,c,d,e)

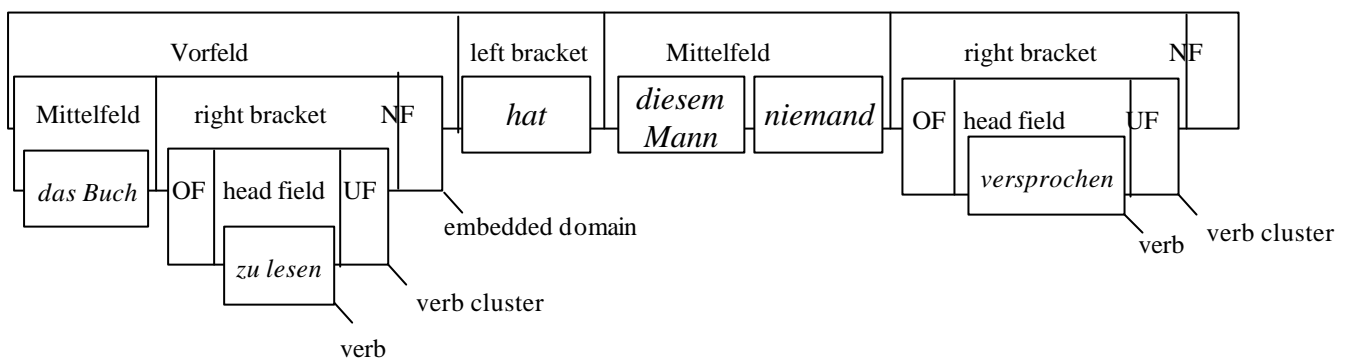
syntactic hierarchy. We think that it is not the case in German. In fact, it makes little sense to form a subdomain for each verb and its dependents. On the contrary, all verbs placed in the same domain lump together all of their dependents in a common pot. In other words, there is no scrambling in German, or more precisely, there is no descriptive advantage in assuming an operation that derives ‘scrambled’ sentences from ‘non-scrambled’ ones.<sup>12</sup>

#### 4.2 Embedding

As we have said, a verb can open an embedded domain, which is placed in one of the major fields. In our reference example, an embedded domain can be opened by *versprochen* ‘promised’ or *zu lesen* ‘to read’ (or even by each of the two verbs). Figure 3 illustrates the case where *zu lesen* opens an embedded domain. We represent domains by ovals with a bold outline. In the situation of Figure 3, *hat* ‘has’ and *versprochen* ‘promised’ occupy the left and right bracket of the main domain and we find three phrases on the same level: *niemand* ‘nobody’, *diesem Mann* ‘to this man’, and the embedded *das Buch zu lesen* ‘to read the book’. The order of these three constituents is free, with one of them having to take the Vorfeld. The embedded domain can go into the Vorfeld (1c), the Nachfeld (1d), or the Mittelfeld (1a,e).

<sup>12</sup> The only remaining advantage of the X-bar based analysis is the alleged universality of the movements, which, however, is quite limited when looked at in detail. The universality of our approach, although neither central to us nor well developed as yet, lies in the embedding of our analysis in the Meaning-Text framework.

Native speakers consider sentences with or without an embedded domain such as (1b) and (1d) as equivalent from the viewpoint of syntactic complexity. Whether or not an embedded domain is created for a subordinated verb depends on the communicative structure. A communicative entity (such as a theme, a rheme, a background, an emphasized group ...) causes its head to open a domain and consequently forms a separated constituent, which can receive its own prosodic contour. Note that the creation of an embedded domain in the Mittelfeld (i.e. the intraposition as in (1e)) is considered as more complex than the creation of an embedded domain in the Vorfeld (i.e. the VP-fronting as in (1c)) or in the Nachfeld (i.e. the extraposition as in (1d)). This contrast could be explained by communicative reasons. The Vorfeld and the Nachfeld are communicatively more typed than the Mittelfeld: The Vorfeld is generally occupied by the



theme or by a prominent rheme (= a focus), while the Nachfeld generally receives a post-rheme, that is non-prominent theme. In some sense, there are few communicative reasons to build an embedded domain and not to place it in the Vorfeld or in the Nachfeld.<sup>13</sup>

<sup>13</sup> An intraposition occurs as a coincidence of different communicative imperatives: If an element carries a topical communicative markup and the Vorfeld is not accessible to this element, then this element can be intraposed. This can happen when the topicalization occurs inside of a sentential complement: Here, the corresponding domain does not possess a Vorfeld and the Vorfeld of the main domain is not accessible from inside the complementizer domain but for wh-words (contrary to English). See example (i) from Kathol 1995:44:

- (i) *Ich glaube, dass [dieses Buch zu lesen] der Professor den Studenten empfohlen hat.*  
 I believe, that [this book to read] the professor to-the students advised has  
 'To read this book is what I believe that the professor has advised the students to do'

In a simple clause, it is necessary that the Vorfeld is already occupied by a more prominent theme and a second, subordinate theme has to be content with the first place in the

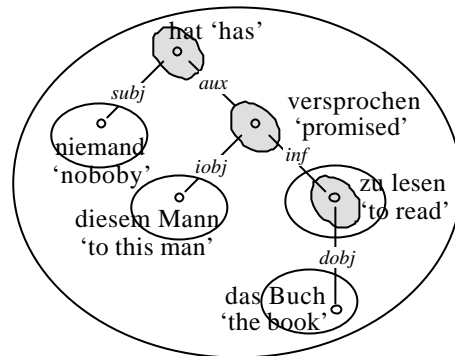


Figure 4: A phrase structure with emancipation for (1f,g)

### 4.3 Emancipation

The dependents of a verb do not have to be in their governor's domain: They can be 'emancipated' and end up in a superior domain.<sup>14</sup> For example, in Fig. 4, the verb *zu lesen* 'to read' has created an embedded domain from which its dependent *das Buch* 'the book' has been emancipated. We have thus four complements to place in the superior domain, allowing more than thirty word orders, among them (1f) and (1g). Among these orders, only those that have *das Buch* or *zu lesen* in the Vorfeld are truly acceptable, i.e. those where embedding and emancipation are communicatively motivated by focus on *das Buch* or *zu lesen*.

An emancipation makes the correspondence between the syntactic dependency tree and the topological structure more complex from a computational viewpoint (whatever the type of computation, by a computer or by a human brain). For this reason, an emancipation must be communicatively well motivated. In our reference example, *das Buch* 'the book' can be emancipated from the domain opened by its governor *zu lesen* 'to read' owing to the fact that *zu lesen* forms a communicative entity without its dependent *das Buch*. For instance, (1f) is appropriate if *zu lesen* is contrasted with another verb as in (3):

---

Mittelfeld as in our example (1e) that could be translated with: *To this man, to read the book, NOBODY has promised that.* (Focus intonation is marked in upper case.)

<sup>14</sup> In cases when an embedded verb has not opened a domain but constructed a verb cluster in the right bracket, as in flat structures, its dependent nominal elements have to be placed in a domain that has been opened by a verb that is not their governor. Nevertheless, we do not consider these cases as emancipations. In some sense, several verbs have agreed in these cases to share the same domain, and consequently their dependents must be considered to be in their governor's domain.

(3) *Zu lesen hat das Buch diesem Mann niemand versprochen, wohl aber zu übersetzen.*

to read has the book to this man nobody promised, well but to translate  
'Nobody promised to the man to READ the book, just to TRANSLATE.

## 5 Extension of the basic German grammar

This section is rather peripheral to the topic of the paper. We provide it, however, to show that our approach based on the topological model allows us to describe key phenomena of German word order. We begin with the word order in the right bracket and pursue with sentential complements, relative clauses, and pied-piping.

### 5.1 *The internal structure of the verb cluster*

As said previously, in the verb cluster a dependent is generally on the left of its governor, giving us the word orders  $V_1, V_2V_1, V_3V_2V_1, V_4V_3V_2V_1, \dots$ . But the reverse order is also possible: The tense auxiliaries HABEN 'have' (past) and WERDEN 'become/will' (future) allow their dependent to be positioned on their right in the verb cluster. This phenomenon, illustrated by the examples in (4), is called *auxiliary flip* or *Oberfeldumstellung* (Bech 1955).<sup>15</sup> In (4a,b,c), the dependent  $V_3$  of the verb  $V_2$ , placed on the right of the auxiliary  $V_1$ , goes again to the left side of its governor  $V_2$ , just as in standard order; we thus obtain the orders  $V_1V_2, V_1V_3V_2, V_1V_4V_3V_2, \dots$ .<sup>16</sup> In example (4c), the role of the fourth verb in the subordination chain is played by a particle rather than a true verb. The resulting order possibilities are identical to the ones for verbal complements.

---

<sup>15</sup> Other names for this phenomenon include *Verb Raising* and *Double-Infinitive Construction*.

<sup>16</sup> It is often argued that the auxiliary flip needs three verbs and that a verb cluster  $V_1V_2$  is impossible. Indeed, the auxiliary flip is only possible with dependent verbs  $V_2$  belonging to a class of modal verb governing an infinitive  $V_3$ . Nevertheless,  $V_3$  does not need to be in the verb cluster and a sentence such as (i) is unproblematic.

- (i) *Das Buch lesen<sub>3</sub> wird<sub>0</sub> er bis morgen wohl haben<sub>1</sub> können<sub>2</sub>*  
the book read becomes he until tomorrow well have can  
'He will have well been able to read the book until tomorrow'

Another possibility is an elliptic construction without apparent dependent of *können* as in the example proposed by Hinrichs and Nakazawa 1994:

- (ii) *... weil er nicht anders hat<sub>1</sub> können<sub>2</sub>*  
... because he not differently has could  
'... because he could not [do it] differently'



- (4) **a.** *Er wird das Buch haben<sub>1</sub> lesen<sub>3</sub> können<sub>2</sub>*  
 he will the book have read can  
 ‘He will have been able to read the book’
- b.** ... *weil er das Auto hätte<sub>1</sub> zu fahren<sub>4</sub> versuchen<sub>3</sub> können<sub>2</sub>*  
 ... because he the car had drive try could  
 ‘... because he could have tried to drive the car’
- c.** ... *weil er das Auto hätte<sub>1</sub> kaputt<sub>4</sub> fahren<sub>3</sub> können<sub>2</sub>.*<sup>17</sup>  
 ... because he the car had broken-drive could  
 ‘... because he could have written off the car’
- d.** *Ich glaube, dass er das Buch wird<sub>1</sub> haben<sub>2</sub> lesen<sub>4</sub> können<sub>3</sub>*  
 I believe that he the book will have read can.  
 ‘I believe that he will have been able to read the book.’

The governed verbs  $V_2$  accepting the inverse order form a closed class including the modal and perception verbs and some other verbs (HELFEN ‘help’, the causative/missive LASSEN ‘make/let’, ...). This class also contains the auxiliary HABEN ‘have’ itself, which suffices to explain the cases of ‘double flip’ as in (4c); we thus obtain the orders  $V_1V_2V_3$ ,  $V_1V_2V_4V_3$ , ...

The dependent  $V_3$  of  $V_2$  can also take the place to the left of the auxiliary  $V_1$ ; we thus obtain the orders  $V_3V_1V_2$ ,  $V_4V_3V_1V_2$ , ... This variant of the auxiliary flip, called the *Zwischenstellung* ‘intermediate position’ or *verbal complex split* (see for example Meurers 1999) and illustrated by the examples in (5), is only accepted by some categories of German speakers. Note that the *Zwischenstellung* can also occur with  $V_4$  rather than  $V_3$  being placed to the left of the auxiliary of the verb cluster, giving us the order  $V_4V_1V_3V_2$  as in (5c). The *Zwischenstellung* is also compatible with the double flip: In this case the verb  $V_4$  can go to the left of the first or the second auxiliary, giving us the orders  $V_4V_1V_2V_3$  and  $V_1V_4V_2V_3$  as in (5d) and (5e).

- (5) **a.** <sup>?</sup>*Ich glaube, dass er das Buch lesen<sub>3</sub> wird<sub>1</sub> können<sub>2</sub>*  
 I believe that he the book read will can.  
 ‘I believe that he will be able to read the book.’
- b.** <sup>?</sup>... *weil ich sonst drei Stunden warten<sub>3</sub> hätte<sub>1</sub> müssen<sub>2</sub>*<sup>18</sup>  
 ... because I otherwise three hours wait had must  
 ‘... because I would have had to wait three hours otherwise’
- c.** <sup>?</sup>... *weil er das Auto kaputt<sub>4</sub> hätte<sub>1</sub> fahren<sub>3</sub> können<sub>2</sub>*  
 ... because he the car broken had drive could  
 ‘... because he could have written off the car’

---

<sup>17</sup> Example taken from Ferret 2002.

<sup>18</sup> Original utterance by Leo Wanner, Paris, April 2002.

d. <sup>?</sup>*Ich glaube, dass er das Buch lesen<sub>4</sub> wird<sub>1</sub> haben<sub>2</sub> können<sub>3</sub>*

I believe that he the book read will have can

'I believe that he will have been able to read the book'

e. <sup>?</sup>*Ich glaube, dass er das Buch wird<sub>1</sub> lesen<sub>4</sub> haben<sub>2</sub> können<sub>3</sub>*

I believe that he the book will read have can

'I believe that he will have been able to read the book'

Note that the dependent of HABEN 'have' must be in the bare infinitive form and not in the past participle form in case of auxiliary flip. This form, called the *Ersatzinfinitiv*, is also possible or even preferable for certain verbs when the auxiliary is in V2 position, as in (6).

(6) *Er hat das Buch lesen können<sup>?</sup> gekonnt*

he has the book read can

'He was able to read the book'

The word order in the verb cluster can be modeled with the same device as the one we used to model the word order in the sentence. We consider that the *verb cluster* is a topological constituent with three fields we call the *Oberfeld*, the *head field*, and the *Unterfeld*. The head field is occupied by the head of the verb cluster, i.e. the verb that creates the verb cluster when occupying the right bracket (or the *Unterfeld* in case of auxiliary flip). Any verbal dependent of the head of the verb cluster can go into the *Oberfeld*, where it opens a topological constituent we call a *verbal box*. Contrarily to the verb cluster, a verbal box has only two fields, an *Oberfeld* and a head field. We need the distinction of verb cluster and verbal box in order to block an auxiliary flip when the auxiliary is in an *Oberfeld* of the right bracket (cf. (7)). To put it simply, once the verb placement down the subordination chain has placed a verb in the *Oberfeld*, no further verb can start an auxiliary flip.

(7) a. *\*Ich glaube, dass Peter das Buch haben<sub>2</sub> lesen<sub>4</sub> können<sub>3</sub> wird<sub>1</sub>*

I believe that Peter the book have read can will

b. *\*Ich glaube, dass Peter das Buch haben<sub>2</sub> wird<sub>1</sub> lesen<sub>4</sub> können<sub>3</sub>*

I believe that Peter the book have want read can

'I believe that Peter will to have read the book.'

The *Zwischenstellung* is modeled by allowing a verb joining an *Oberfeld* to pass through verbal boxes and verb clusters and landing in the *Oberfeld* of a verb cluster opened by an auxiliary and vacant because the dependent of the auxiliary has gone in the *Unterfeld*. As we see, the *Zwischenstellung* is very natural and easy to describe with our grammar. The *Zwischenstellung* can easily be permitted or not according to whether we allow a verb placed in the *Oberfeld* not only to go to the *Oberfeld* of its governor, but to the *Oberfeld* of the whole verb cluster. This nicely explains the acceptance variations among German speakers.

In related languages like Dutch or Swiss-German, which have the same topological structure, the standard order in the right bracket is somewhat similar to the German Oberfeldumstellung, that is, a bare infinitive tends to be on the right of its governor (*te*-infinitives, the Dutch equivalent of German *zu*-infinitives, cannot be embedded in the right bracket and must absolutely open an embedded domain). The resulting order gives rise to cross serial dependencies (Evers 1975, Bresnan *et al.* 1982).<sup>19</sup> Such constructions have often been studied for their supposed complexity. Our German grammar can easily be adapted to Dutch (keeping in mind that we do not describe the order of the Mittelfeld) using exactly the same topological constituents: A verb in the right bracket or in an Unterfeld still opens a verb cluster with the same three fields. Contrarily to German, the rule allowing a verb to go in an Oberfeld is restrained (it seems that only past participles can go in the Oberfeld), while the rule allowing a verb to go in the Unterfeld is relaxed, the Oberfeldumstellung being seemingly allowed to the whole class of verbs governing a bare infinitive (contrarily to German where it is restricted to auxiliaries). Note that a particle (= separable prefix) must go in the Oberfeld. But as in German, for some speakers the Oberfeld of all the constituents containing the head of the particle are available, giving us cases of Zwischenstellung, illustrated by (8) (borrowed from Bouma and van Noord 1998, example 58)

- (8) **a.** ... *dat Jan Marie zou<sub>1</sub> hebben<sub>2</sub> aan<sub>4</sub> gesproken<sub>3</sub>*  
 ... that Jan Marie would have on spoken  
 ‘... that Jan would have spoken to Marie’
- b.** ... *dat Jan Marie zou<sub>1</sub> aan<sub>4</sub> hebben<sub>2</sub> gesproken<sub>3</sub>*
- c.** ... *dat Jan Marie aan<sub>4</sub> zou<sub>1</sub> hebben<sub>2</sub> gesproken<sub>3</sub>*
- d.** \*... *dat Jan Marie zou<sub>1</sub> hebben<sub>2</sub> gesproken<sub>3</sub> aan<sub>4</sub>*

We hope that we have convinced the reader that the topological model applies to the description of the word order in the right bracket in a remarkably elegant manner. The extension of the topological model to the internal structure of the verb cluster is, as far as we know, quite new; previous approaches based on the topological model, such as Kathol 1995, 2000, used different devices to sort out the problem of the word order in the right bracket.

---

<sup>19</sup> As we have said previously, the order in the Dutch Mittelfeld is syntactically constrained and the constituents must appear in their hierarchical order: A constituent  $C_1$  depending on  $V_1$  must appear before a constituent  $C_2$  depending on  $V_2$  and so on. Due to the order in the right bracket we obtain orders such as  $C_1C_2\dots C_nV_1V_2\dots V_n$ , as in (i) (borrowed to Bresnan *et al.* 1982).

- (i) ...*dat Jan<sub>1</sub> Piet<sub>2</sub> Marie<sub>3</sub> de kinderen<sub>4</sub> zag<sub>1</sub> helpen<sub>2</sub> laten<sub>3</sub> zwemmen<sub>4</sub>*  
 ...that Jan Piet Marie the children saw help let swim  
 ‘...that Jan saw Piet help Marie to make the children swim’

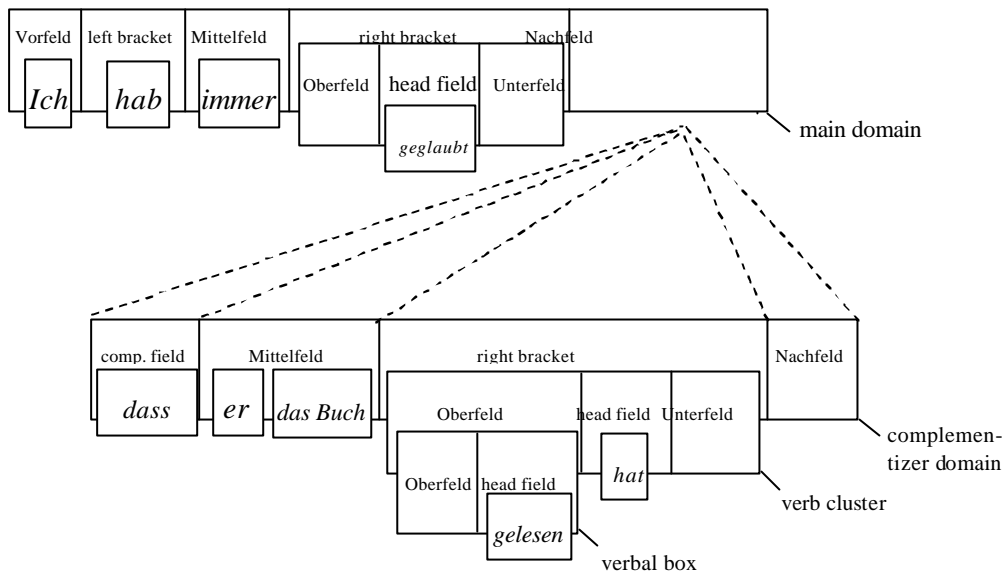


Figure 5: A complementizer domain in the Nachfeld of a main domain

## 5.2 Sentential Complements

In sentential complements headed by a complementizer<sup>20</sup>, the complementizer is in the first position and the main verb occupies the final position (which can be followed, exactly as in the main clause, by extraposed constituents). Moreover, we consider, following Tesnière 1959, Mel'cuk 1988 and the recent versions of the X-bar Syntax, that the complementizer is the head of the completive clause. Consequently, we treat the order in the completive clause by considering that the complementizer opens a domain we call the complementizer domain composed of a sequence of four fields: the left bracket, the Mittelfeld, the right bracket and the Nachfeld. The left bracket of the complementizer domain is also called the complementizer field (Kathol 1995). Exactly as in the main domain, the left bracket is the head field, which is occupied by the head of the completive clause, the complementizer. The dependent of the complementizer – the finite main verb of the clause – must go in the right bracket where it opens, as usual, a verb clus-

<sup>20</sup> There are also cases of sentential complements without complementizer. Their behavior is easily predictable from our analysis: The domain is opened by the highest verb that goes into the complementizer field, and its dependents are as usually placed into the right bracket. See Boethius' famous sentence in (i):

- (i) *Hättest du geschwiegen, wärest du ein Philosoph geblieben.*  
 had you been-silent, would you a philosopher remained.  
 'If you had kept quiet, you would have remained a philosopher.'

ter.<sup>21</sup> See (9) and the corresponding topological phrase structure in Fig. 6 as an example.

- (9) *Ich habe immer gedacht, dass er das Buch gelesen hat.*  
 I have always thought, that he the book read has  
 ‘I always thought he had read the book.’

### 5.3 Relatives and pied-piping

The relative pronoun takes the first position of the relative clause, modulo the fact that it can take other elements along. This phenomenon, called *pied-piping*, is illustrated by example (10). German differs from English and Romance languages in that even verbs can be brought along by the relative pronoun, as in (10b,c).<sup>22</sup>

- (10)a. *Der Mann* [[von dem] [Maria] [geküsst wird]] *liebt sie.*  
 The man [[by whom] [Maria] [kissed is]] loves her
- b. *Das war eine wichtige Einnahmequelle, [[die zu erhalten] [sich] [die EU] [verpflichtet hat]].*  
 This was an important source-of-income, [[which to conserve] [itself] [the EU] [committed has]]  
 ‘This was an important source-of-income, that the EU obliged itself to conserve.’
- c. *Alles dies sind Fragmente des gigantischen postsowjetischen Syndroms, [[von dem sich zu kurieren] [bislang] [nur sehr wenigen] [ge-*

---

<sup>21</sup> The finite verb depending on the complementizer cannot open a domain. The case of a word that cannot open a domain is rather rare. In some sense, the complementizer and the finite verb are inseparably linked, as the determiner and the noun are, which must also always be in the same domain. The parallelism of complementizer-verb on one hand and determiner-noun on the other hand has been considered by many linguists from Tesnière 1959 (with its theory of ‘translation’) to Abney 1987 (with the DP-hypothesis in the X-bar framework).

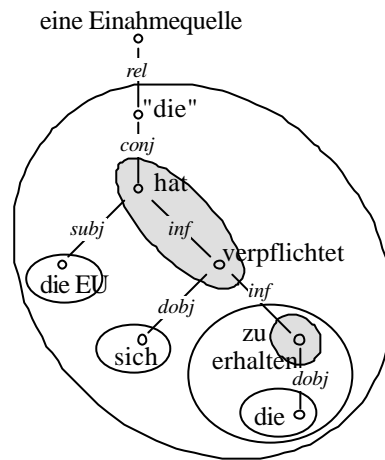
<sup>22</sup> The relative clauses (10b,c) can receive another description where the relative pronoun occupies the first position alone, the infinitive being intraposed and the relative pronoun being emancipated from the domain it opens. As said before, the intraposed construction is very marked and rather improbable in this context. Moreover, all pronouns, like *sich* ‘itself’ in (10b), tend to be in the left of the Mittelfeld, which makes it even more improbable for *zu erhalten* ‘to conserve’ to be in the Mittelfeld. The reduced acceptability of (i), compared to (10b), can be taken as a proof:

- (i) ??*Das war eine wichtige Einnahmequelle, [[die] [sich] [zu erhalten] [die EU] [verpflichtet hat]]*

*lungen ist*]].<sup>23</sup>

all this are fragments of-the gigantic postsoviet syndrome,  
 [[from which oneself to cure] [yet] [only very few] [succeeded is]]  
 ‘All this are fragments of the gigantic post soviet syndrome which only few  
 have yet succeeded to cure.’

Before we discuss the topological structure of relative clauses, we will discuss their syntactic representation. Following Tesnière (1959) and numerous analyses that have since corroborated his analysis (see Kahane *to appear*), we assume that the relative pronoun plays a double syntactic role:



**Figure 6: The dependency tree and the topological structure of the relative clause in (10b)**

- On the one hand, it has a pronominal role in the relative clause where it fills a syntactic position.
- On the other hand, it plays the role of a complementizer allowing a sentence to modify a noun.

For this reason, we attribute to the relative pronoun a double position: as a complementizer it is the head of the relative clause and it therefore depends directly on the antecedent noun and it governs the main verb of the relative clause; as a pronoun, it takes its usual position in the relative clause.

It is now possible to give the word order rules for relative clauses. The complementizing part of the relative pronoun opens a complementizer domain, exactly as does the complementizer of a sentential complement. The main verb that depends on it joins the right bracket. The other rules are identical to those for other domains, with the group containing the pronominal part of the relative pronoun having to join the other part of the pro-

<sup>23</sup> Taz journal Nr. 6636, 28.12.2001, page 4, article *Die Beerdigung ist nicht zu Ende* by Michail Ryklin, Berlin.

noun in the complementizer field. In a sense, the complementizer field acts like the fusion of the Vorfeld and the left bracket of the main domain: The complementizing part of the pronoun, being the root of the dependency tree of the relative clause, takes the left bracket (just like the top node of the whole sentence in the main domain), while the pronominal part of the relative pronoun takes the Vorfeld. The fact that the pronoun is one word requires the fusion of the two parts and hence of the two fields into one. Note that verbal pied-piping is very easy to explain in this analysis. It is just an embedding of a verb in the complementizer field: Just like the Vorfeld, the complementizer field can be occupied by a non-verbal phrase or by a verb creating an embedded domain.

## 6 Characterizing the notion of topological phrase

In this section, we will try to be more precise about the status of our notion of topological phrase and attempt a first characterization.

In a sense, the simple fact that we have succeeded in describing a large spectrum of German word order phenomena in a simpler way than previous descriptions could be sufficient to validate our approach and consequently our notion of topological phrase. We want to show now that topological phrases are not only useful gears in our mechanism, but can actually be observed in the sentence and correspond to entities present in the consciousness of the speaker, namely communicative groupings and prosodic constituents.

### 6.1 *Syntactic tests*

Let us first recall that many tests have been proposed to characterize the phrases in the framework of phrase structure grammar (for the precursors see Bloomfield 1933, Wells 1947, Harris 1951; for recent surveys and critics of these tests see Bonami 1999 and Abeillé 1991). These tests allow us to determine syntactic constituents, a notion that is different from the notion of topological phrase we investigate here. Nevertheless, it is instructive to take a quick look at these tests.

Utterance tests: A group of words is a potential syntactic constituent if it can constitute a whole utterance, for instance if it can be an answer to a question (*What does Peter try? To read the book*)

Commutation tests: A group of words is a potential syntactic constituent if it can commute with a single word, for instance a pronoun (*Peter tries to read the book → Peter tries that*)

Position tests: A group of words is a potential syntactic constituent if it can appear in a syntactic position that is known to be occupied by a single constituent. For instance, in English, topicalization and clefting constitute good tests (*To read the book Peter tries; It is to read the book what Peter tries*). In German, the occupation of the Vorfeld provides such a test, as illustrated by (11a). Nevertheless, the test is weakened by the

possibility of emancipation: In (11b), the group *zu lesen* ‘to read’ can occupy the Vorfeld but it will not be recognized as a syntactic constituent since its syntactic dependent *das Buch* did not co-occur in the Vorfeld, but only later in the sentence.

- (11)a. Das Buch zu lesen versucht Peter  
 the book to read tries Peter  
 b. Zu lesen versucht Peter das Buch  
 to read tries Peter the book

We do not reject the usefulness of the previous tests. In our perspective they will be used to justify our choices in the syntactic representation that we prefer to express with a dependency tree. An X-bar phrase structure tree can more or less equivalently express the same syntactic structure. The syntactic tests proposed for the determination of syntactic constituents (with the addition of the tests proposed for the determination of the *head* of a syntactic constituent) can directly be adapted to determine precisely the syntactic dependencies (see Mel’cuk 1988 for equivalent tests for the determination of dependency).

Our topological phrases generally do not pass the syntactic tests above. Let us take the example of the verb cluster. It can not in general be a complete utterance, it does not generally commute with a single word and it cannot occupy another position than the right bracket of a domain (and inversely only a verb cluster can occupy this position). It may be worthwhile to note that the position of the verb cluster can be occupied by a single verb and that, in some sense, a verb cluster behaves as a single verb: The dependents of the different verbs of the verb cluster behave as if they were the dependent of one word, their relative order being free. Nevertheless, in a given sentence such as for instance (12), it is hard to replace the verb cluster by a single verb (no German verb subcategorizes two dative complements):

- (12)... *weil dem Mann seinem Sohn niemand zu helfen anbietet*  
 ... because to-the man to-his son nobody to help offers  
 ‘... because nobody offers the man to help his son ’

If the verb cluster is clearly not a syntactic constituent, most of the topological phrases coincide with syntactic constituents. This is true for all domains. A domain is the largest topological phrase a word can open and this corresponds to its maximal projection. (The maximal projection of a word is the projection of the whole subtree it roots in the syntactic dependency tree.) This last property is very important as it characterizes the mutual constraints between the syntactic representation and the topological phrase structure. We think that this strong relation between syntactic constituent and topological phrase could explain why many theories do not distinguish them. These theories start out with the idea that every element basically projects maximally and consequently must use complex devices (such as movements) when they are confronted to mismatches between syntactic constituents and topological phrases. Note that in our approach, the domain is the largest topological constituent a word can open, which means that this



constituent can accommodate all the dependents of the word, but it certainly does not mean that this accommodation is obligatory, due to possible emancipations.

We will pursue our attempt of characterization of the topological phrases in the direction of prosody. Prosody has often been put forward to determine the syntactic constituents, although some mismatches between syntactic constituents and prosodic units have long been revealed (Chomsky and Halle 1968, Martin 1981, Hirst and Di Cristo 1998). We will now argue that prosodic tests are more suitable to determine topological phrase.

## 6.2 *Prosodic tests*

A particularity of our approach is that a syntactically non-ambiguous sentence can have several topological phrase structures. From the viewpoint of phrase structure grammars, these would be considered as spurious ambiguities and not accepted.<sup>24</sup> We will show that these structural ambiguities are justified both from a communicative standpoint and from a prosodic standpoint.

For the word order presented in (1a) we obtain three different topological structures (but of course only one syntactic dependency tree) schematized in (13). See Figure 7 for the detailed topological structures.<sup>25</sup>

- (13) **a.** [*Die Männer*] [*haben*] [*dieser Frau*] [*den Roman*] [*zu lesen versprochen*]  
**b.** [*Die Männer*] [*haben*] [*dieser Frau*] [*den Roman zu lesen*] [*versprochen*]  
**c.** [*Die Männer*] [*haben*] [*dieser Frau*] [*den Roman*] [*zu lesen*] [*versprochen*]  
the men have to-this woman the novel to read promised  
'The men promised to this woman to read the novel'

---

<sup>24</sup> See for example the discussion in Müller 1999, section 17.5.

<sup>25</sup> The dependency structures of (1a) and (13) are identical except for lexical elements.

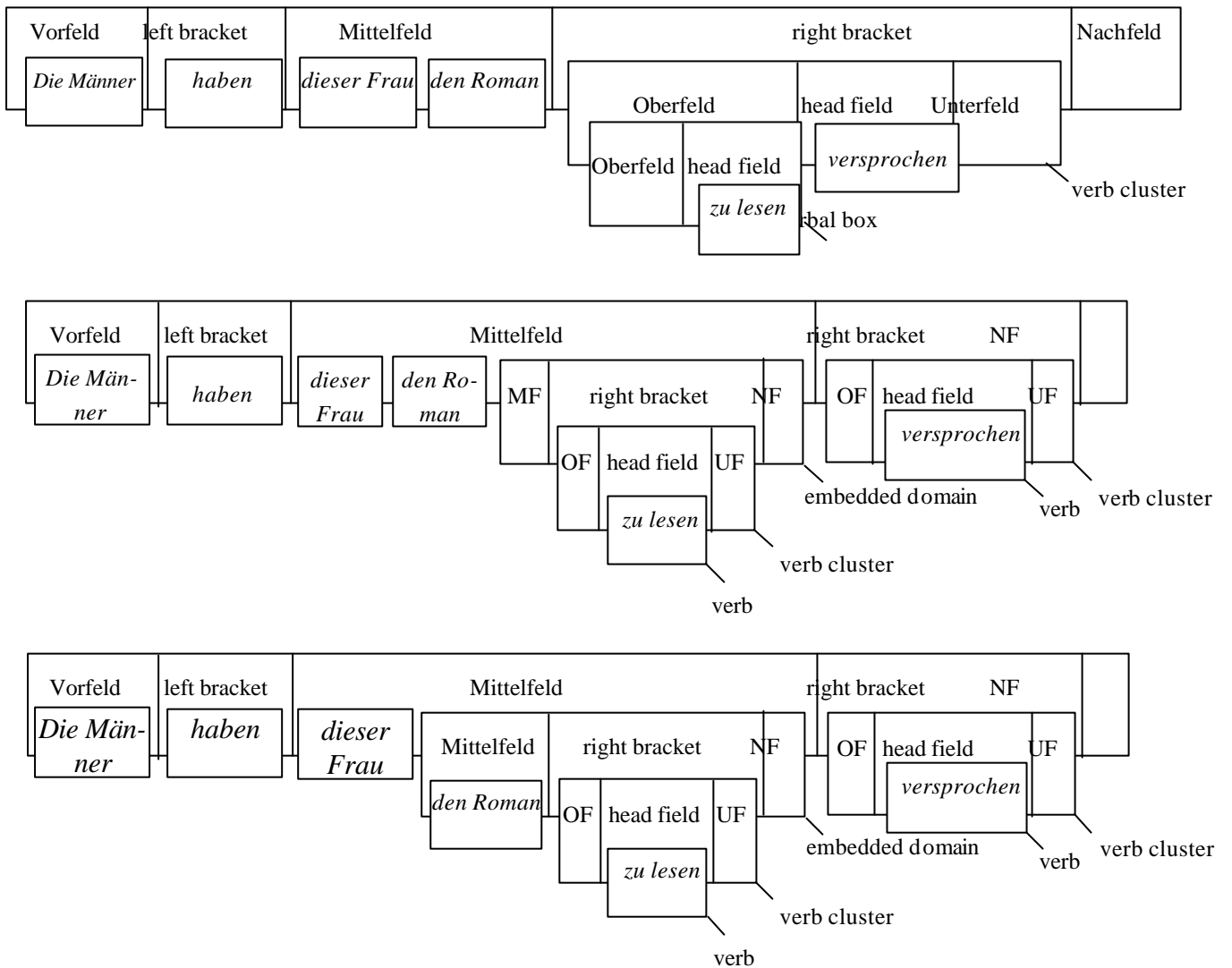


Figure 7: Three different topological structures for the same word orders

The first structure is the flat structure, where the verbs *versprochen* and *zu lesen* go in the right bracket. The second structure illustrates the intraposition: the *verb zu lesen* opens an embedded domain in the Mittelfeld. The third structure involves the intraposition of *zu lesen* with an emancipation of *den Roman*, which is thus placed beside *zu lesen* in the Mittelfeld of the main domain.

This structural ambiguity corresponds, we believe, to a semantic ambiguity of communicative type. In the three structures, *die Männer* 'the men' is in the Vorfeld. Sentences with *die Männer* 'the men' in the Vorfeld can be natural answers to a question of type *Was ist mit den Männern?* 'What is happening with the men?'. In this case *die Männer* is the theme. The constituent in the Vorfeld can also express a focus (or prominent rheme), prosodically clearly distinct from the theme prosody (see Choi 1999, Gibbon

1998, and Büring 1997 for details). The focus case would rather be translated with a cleft construction (like *It is the men that promised her to read the novel*). In (13a), if *die Männer* is the theme, the rest of the sentence will be a neutral rheme, while in (13b) the fact of reading the novel is marked. To understand this marking, we compare (13b) with (14):

(14) [*Den Roman zu lesen*] [*haben*] [*dieser Frau*] [*die Männer*] [*versprochen*]

In (14), the same dichotomy as in (13) holds for the constituent in the Vorfeld: This constituent can either be the theme (corresponding to ‘To read the novel, the men have promised that to this woman.’) or prominent rheme (‘It is to read the novel what the men promised to this woman.’). The case of sentence (13b) is more complex: such a case could arise in a situation where *den Roman zu lesen* appears as a distinct communicative unit, either because it is itself contrasted with another action *ihre Zimmer aufzuräumen* ‘clean up their rooms’ as in (15a), or the governor *versprochen* ‘promised’ is contrasted with another control verb as in (15b), thus disallowing the formation of a verb cluster.

(15)a. [*Die Männer*] [*haben*] [*dieser Frau*] [*den Roman zu lesen*] [*versprochen*]  
und [*die Kinder*] [*ihre Zimmer aufzuräumen*]

the men have to-this woman the novel to read promised and the children  
their rooms up-to-clean

‘The men promised to this woman to read the novel and the children, to  
clean up their rooms’

b. [*Die Männer*] [*haben*] [*dieser Frau*] [*den Roman zu lesen*] [*ver-*  
*sprochen*] und [*die Kinder*] [*gedroht*]

the men have to-this woman the novel to read promised and the children  
threatened

‘The men promised to this woman to read the novel and the children  
threatened it.’

The structure (13c) is hard to motivate communicatively. The embedded domain opened by *zu lesen* and its emancipated complement *den Roman* have been created, but it remains unclear why, since they are not used for the clear expression of the sentence’s communicative structure, which could be accomplished by fronting or extraposition.<sup>26</sup> In consequence, we can predict that a structure such as (13c) will hardly ever be produced, which seems to be the case.

---

<sup>26</sup> These isolated elements ([*den Roman*] and [*zu lesen*]) could remain side by side in the Mittelfeld in an extremely complex communicative structure where the fronting positions are unavailable or already occupied by communicatively even more prominent elements (cf. footnote 13). Sentence (i) has as its only topological structure an intraposition with emancipation into the Mittelfeld while Vorfeld and Nachfeld seem to remain accessible, just as the topological structure in sentence (13c). The unavailability of an

We will now show the close connection that exists between prosody and topology. The verb cluster *zu lesen versprochen* ‘to read promised’ occupying the right bracket of the main domain can be seen in Figure 8.<sup>27</sup> This verb cluster is characterized by only one accent (on the first syllable of the radical of the first verb without its unaccented prefix) followed by a regular fall of the melodic contour. Equally in Figure 9, the right bracket is reduced to *versprochen* ‘promised’, that now carries an intensity accent on the first syllable of the radical (*spro*) and a regular fall of the melodic contour. We also discern the embedded domain *den Roman zu lesen* ‘the novel to read’, set off by a fall of the fundamental frequency and an accentuation on its right bracket occupied by *zu lesen* ‘to read’. Moreover, the melodic contour remains flat after the initial pitch accent, contrasting with the example in Figure 8 where *zu lesen* forms a verb cluster with the verb following it (*versprochen*).

The example in Figure 10 ends just as the example in Figure 8 on the string *den Roman zu lesen versprochen*. Yet, the prosodic representation bears a closer resemblance with the contour of Figure 9: *den Roman zu lesen* possesses exactly the same melodic contour as in Figure 9, while *versprochen* presents a pitch accent and a falling contour, which makes us want to say that it occupies the right bracket alone.

Figures 8 and 9 have been obtained in reading without any specific indication. Figure 10 however emerged in a situation when *den Roman zu lesen* was previously introduced in the discourse. Similar results have been obtained in spontaneous speech as answers to questions backing this grouping like for example *Was haben die Männer dieser Frau versprochen?* ‘What did the men promise to the woman?’

---

easy communicative context explains why (i) is sometimes considered ungrammatical. We prefer to judge the sentences (i) and (13c) as topologically correct, but communicatively unmotivated.

- (i) ?[Die Männer] [haben] [zu lesen] [dieser Frau] [den Roman] [versprochen]  
the men have to read to-this woman the novel promised  
‘The men promised to this woman to read the novel’

<sup>27</sup> Figure 8, 9, and 10 present each the fundamental frequency contour, the aligned text, the intensity contour and the signal.

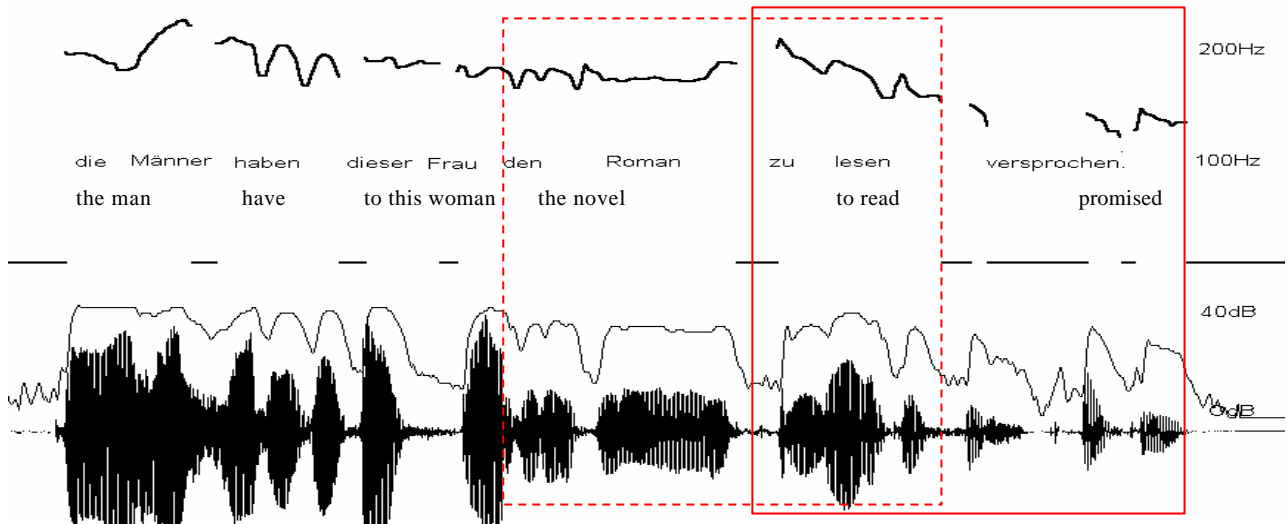


Figure 8. Verb cluster (prosody of (13a))

— correct segmentation  
 - - - incorrect segmentation

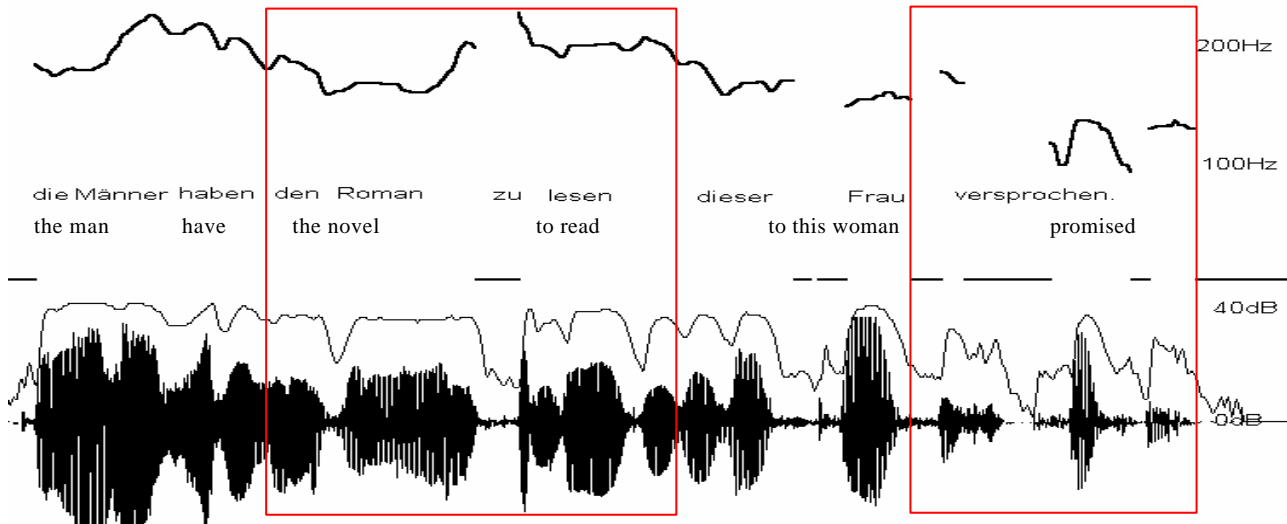


Figure 9. Obligatory embedding

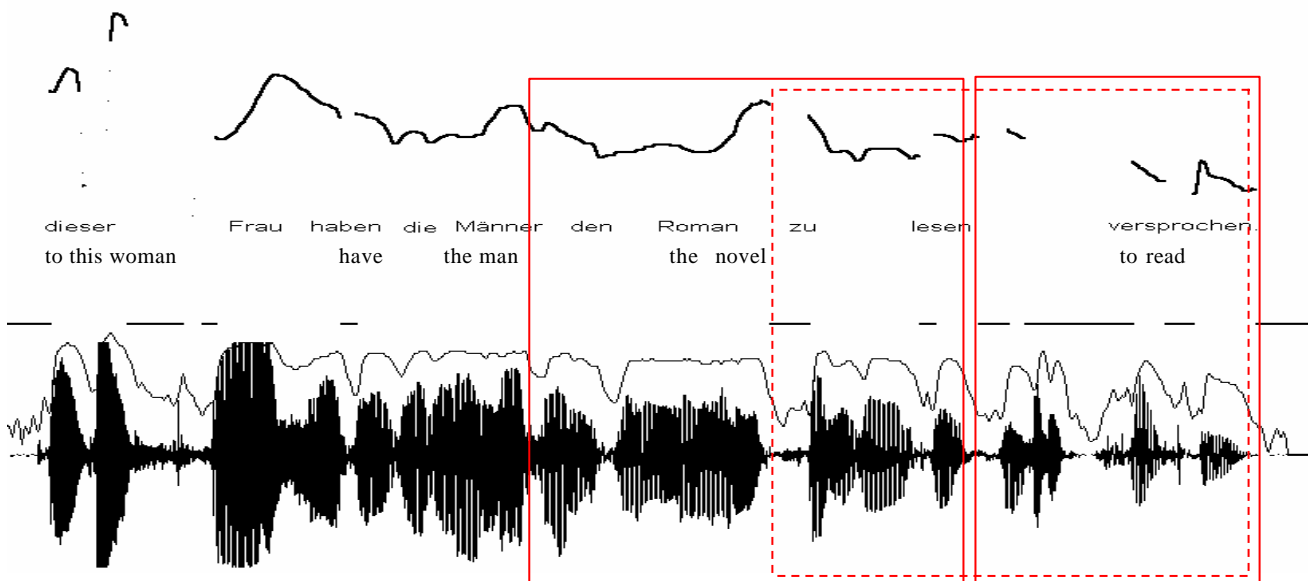


Figure 10. Supposed embedding (prosody of (13b))

The existence of the structural ambiguity is also confirmed by the contrast between full infinitives (with *zu*) and bare infinitives (without *zu*): Bare infinitives cannot form an embedded domain outside of the Vorfeld. Consequently, there are two different contours for (16a) (with or without detachment of *den Roman* ‘the book’ from *zu lesen* ‘to read’), whereas only one prosodic contour without detachment is permitted for (16b), although (16a) and (16b) have isomorphic dependency trees. Evidence comes also from the written form, where a comma is recommended for (16a) (i.e. preference for the embedded structure), whereas the comma is not allowed for (16b).

(16)a. *Niemand versucht(,) den Roman zu lesen*

‘Nobody tries to read the novel.’

b. *Niemand will den Roman lesen*

‘Nobody wants to read the novel.’

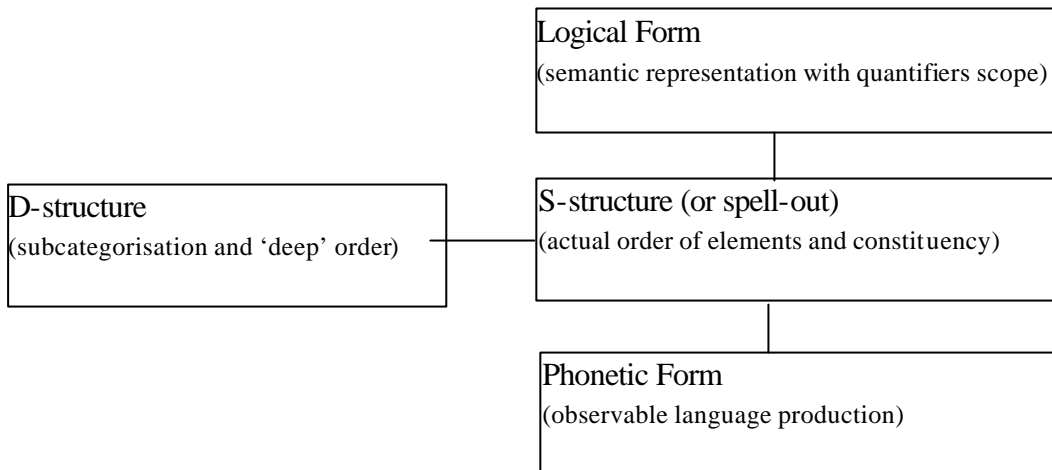
## 7 The place of the phrase structure in the linguistic model

In the first part of this article, we contrasted our view on the phrase with the one of so-called phrase structure grammars of the Chomskyan tradition. In its more recent forms, the Chomskyan framework takes linguistic description as a pairing of meanings and texts (the conceptual-intentional level and the articulatory-perceptual level in Chomsky 1995’s terms). On the assumption of the descendants of the Extended Standard Theory “each language will determine a set of pairs (p,?) [p drawn from Phonetic Form (sic !) and ? from Logical Form] as its formal representations of sound and meaning, insofar as these are determined by the language itself” (Chomsky 1995: 169). This is a point of convergence between Chomsky’s and Mel’cuk’s approaches. This pairing is the founding and name-giving idea of the Meaning-Text Theory: Postulate 1 of the theory “means that a natural language is viewed as a logical device that establishes the correspondence between the infinite set of all possible meanings and the infinite set of all possible texts and vice versa. For a given meaning, this device must ideally produce all the texts that, in the judgment of native speakers, correctly express this meaning, thus simulating SPEAKING; from a given text, the device must extract all the meanings that, according to native speakers, can be correctly expressed by the text, thus simulating SPEECH UNDERSTANDING” (Mel’cuk 1988: 44).

Nevertheless, the two approaches differ in the architecture put in between the two so-called interface levels of the linguistic models, and in particular, they hold different views on the status of the phrase structure.

In Chomskyan models, phrase structure with its theoretical foundation, the X-bar theory, is the central structural description. “The computational system takes representations of a given form and modifies them. Accordingly, Universal Grammar must provide means to present an array of items from the lexicon on a form accessible to the computational system. We may take this form to be some version of X-bar theory. The con-

cepts of X-bar theory are therefore fundamental. In a minimalist theory, the crucial properties and relations will be stated in the simple and elementary terms of X-bar theory” (Chomsky 1995: 172, emphasis is ours). The X-bar idea is taken for granted and not further justified than by its wide usage. Neither the concept itself nor its place in the linguistic model are called into question.



**Figure 11: The classical diagram for the architecture of Chomskyan models**

The phrase structure has the burden to testify the complete correspondence between meaning and text. In Figure 11, we recall the classical diagram for the architecture of Chomskyan models (Chomsky 1965, 1995). From the D-structure (already an X-bar phrase structure) the S-structure is derived (another X-bar phrase structure with empty categories). The position of words in the S-structure is close but not necessarily identical to the final position appearing in the Phonetic Form (besides phonological elisions, some approaches also have clitic movement happen between the surface structure and the phonological structure). Similarly, the S-structure allows a direct derivation of the Logical Form.

Thus, the S-structure encodes the order of elements, their former places (traces) in the D-structure, the argument structure between different nodes, and their scope relations.

The choices involved in the formation of the S-structure are guided by the idea that the implied derivations should be as simple as possible. Thus, an S-structure has to be as close as possible to the structures encoding information as different as scope relation, linear order of elements, and subcategorization frames. It is clear that this can only be achieved by an extremely complex S-structure, where empty nodes and co-indexations are abundant.

From a Meaning-Text viewpoint, a phrase structure such as the S-structure contains both the representations of meaning (LF) and of text (PF), as well as all the intermediate representations – and also the correspondences between these representations. This

means that it does not only contain the representations themselves, but it also shows the interrelations between them.

This can be represented as in Figure 12: Each meaning-text pairing  $(\langle, p)$  is realized via an S-structure, that is itself a projection of a D-structure. We represent the S-structures as ovals between meaning representations (LF) and texts (PF), represented by gray rectangles. Each S-structure is linked to a D-structure in a third dimension.

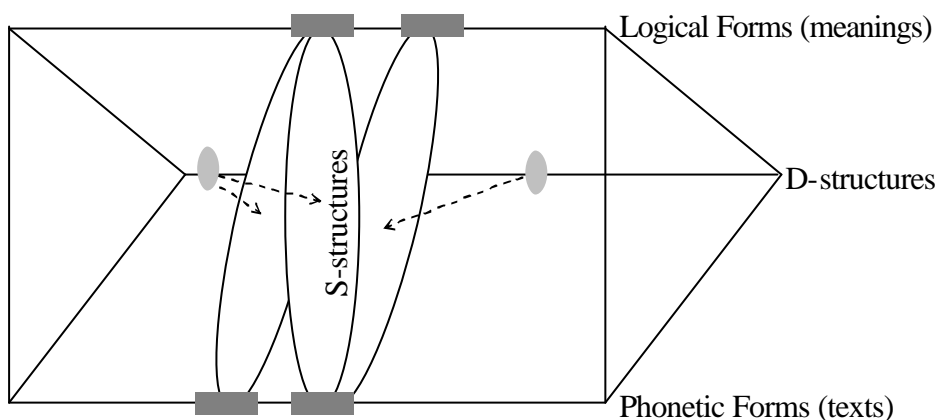


Figure 12: A more expressive view on the architecture of the Chomskyan model

In our approach based on the Meaning-Text framework, contrary to the Chomskyan architecture, the correspondence between meanings and texts passes through intermediate representations, among them the (surface) syntactic representation and the topological representation.

In the Meaning-Text model it becomes clear that the same meaning can be conveyed by different lexical choices (an early choice in the language generation process), by different syntactic constructions, by different word orders and groupings, and also by different intonation patterns (a late choice in the language generation process). Moreover, the communicative structure of the message is present and influences the different choices coming up on the way from semantics to intonation.

We illustrate the Meaning-Text architecture in Figure 13: We depict again the different representations by gray rectangles and their correspondences by ovals. Note that these correspondences are not considered by themselves, and do not receive their own representation (for an alternative view inside the Meaning-Text framework, ascribing structures to correspondences, see Kahane 2002).

It is clear that the information that the Chomskyan tradition wants to put into phrase structure is very different from the task we give to (topological) phrase structure. The topological representation is just an intermediate step on the way from meaning to text representing the way words in the surface string are grouped together. Conversely, a node in a Chomskyan phrase structure tree is conditioned by various goals: The surface



grouping of words, but also syntactic and semantic considerations including even quantifier scope.<sup>28</sup>

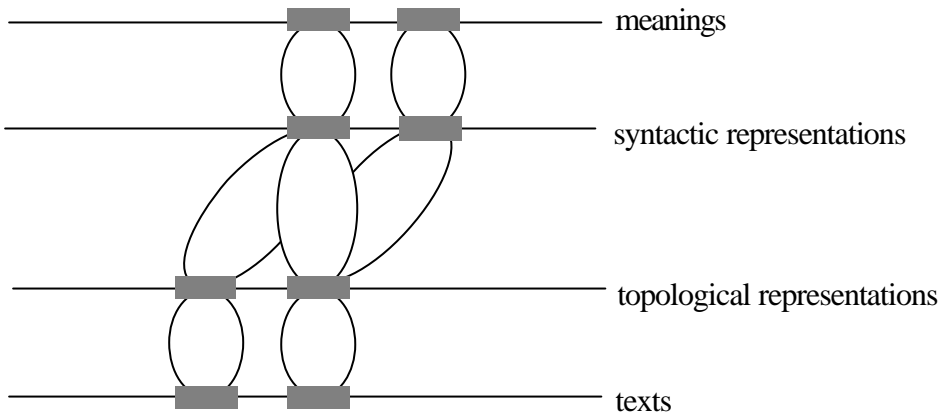


Figure 13: Diagram for the architecture of a Meaning-Text model

## 8 Conclusion

We subscribe to the idea underlying science that separable information should be represented separately. The application of this principle in linguistics has been vigorously defended by Igor Mel'cuk through the Meaning-Text Theory. It allows for a clearer vision of the various parameters intervening in the correspondence between meanings and texts. We defend the idea that phrases, that is, ordered groupings of words, must be separated from the syntactic dependency and that these two notions must appear in two different structures, the goal of syntax being to interface these two levels of representation.

We have presented an interface between syntactic structures and topological phrase structures for German. Our syntactic structures are not linearized. They include the dependency structure and the communicative grouping, which hold a direct link to the semantic level of representation. The topological phrase structures, on the other hand, are set up when linearizing. This allows us to construct all possible word orders for the given syntactic structure, and our phrases can be linked directly to prosodic units.

The simplicity and power of our German grammar stems from the fact that we started off with a syntactic representation whose well-formedness is taken care of by a different

---

<sup>28</sup> Following Mel'cuk 2001, we think that the quantifier scope is closely linked to the communicative structure, which is important for the choices involved in building the topological phrase structure. Thus, quantifier scope indirectly influences the order of the sentence.

module (the semantic-syntax interface). For example, the root of the syntactic dependency tree—the syntactically highest verb of a sentence—can be introduced by very different semantic configurations according to the fact that it is a full verb (*Peter liebt Maria* ‘Peter loves Maria’), an auxiliary (*Peter hat Maria geliebt* ‘Peter loved Maria’), a raising verb (*Peter scheint Maria zu lieben* ‘Peter seems to love Maria’), a light verb (*Peter stellt Maria eine Frage* ‘Peter asks Maria a question’), a part of an idiom (*Peter beißt ins Gras* ‘Peter kicks the bucket’). Whatever the case, the verb will go into the left bracket of the main domain and its complements will behave in the same manner. Consequently, only one simple rule is needed for the syntactically highest verb in the syntactic-topological interface. If however, we wanted to relate directly the semantic representations with the word order level, our rules would be considerably more complicated and the word order rules would be difficult to distinguish from semantic and syntactic considerations. The major argument for topological phrase remains the economy of the system.

Let us recall that our strict separation of subcategorization and phrase structure allows for the same lexical unit to open very different phrases: A verb placed in the right bracket of its governor’s domain opens a reduced phrase—the verb cluster—that can only accommodate one other verb (and by recursivity a string of verbs), whereas a verb placed in a major field opens an embedded domain that can accommodate all of its dependents.

In conclusion, we advocate a remodeling of phrase structure. Phrase structure is the result of the combination of communicative structure and syntactic dependency (itself linked to subcategorization), in accordance to language internal rules, but syntactic dependency should not itself be part of phrase structure (as it the case in phrase structure grammars based on the X-bar theory). Phrases only intervene when word order is at play, at an intermediate level of the utterance’s representation between the syntactic representation and the phonological representation.

## References

- Abeillé, A. 1991. *Une grammaire lexicalisée d'arbres adjoints pour le français*. PhD thesis, Université Paris 7. Updated version to appear.
- Abeillé, A. 1996-97. “Fonction objet ou position objet (1<sup>ère</sup> et 2<sup>e</sup> parties)”. *Gré des langues*, 11, 8-29 ; 12, 8-33.
- Abney, S. 1987. *The English Noun Phrase in its Sentential Aspect*. PhD thesis, MIT, Cambridge.
- Bech, G. 1955. *Studien über das deutsche verbum infinitum*. Copenhagen. Reprinted Niemeyer, Tübingen, 1983.

- Bloomfield, L. 1933. *Language*, New York.
- Bonami, O. 1999. *Les constructions du verbe : le cas des groupes prépositionnels argumentaux*. PhD thesis, Université Paris 7.
- Bouma, G. & G. van Noord. 1998. "Word Order Constraints on Verb Clusters in German and Dutch", in E. Hinrichs, T. Nakazawa, and A. Kathol (eds.). *Complex Predicates in Nonderivational Syntax*. Academic Press, New York.
- Bresnan, J., R. Kaplan, S. Peters & A. Zaenen. 1982. "Cross-serial Dependencies in Dutch", *Linguistic Inquiry*, 13:4, 613-635.
- Büring, D. 1997. *The 59th Street Bridge Accent: On the Meaning of Topic and Focus*, Routledge, London.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*, MIT Press, Cambridge.
- Chomsky, N. 1995. *The Minimalist Program*, MIT Press, Cambridge.
- Chomsky, N. & M. Halle. 1968. *The sound pattern of English*. Harper & Row, New York. Reprinted MIT Press, Cambridge, 1991.
- Choi, H.-W. 1999. *Optimizing Structure in Context – Scrambling and Information Structure*. CSLI, Stanford.
- Di Cristo, A. & D. Hirst (eds.). 1998. *Intonation systems: A survey of twenty languages*. Cambridge University Press, Cambridge.
- Drach, E. 1937. *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt.
- Donohue, C. & I. Sag. 1999. "Domains in Warlpiri", *Proceedings of HPSG99*, Edinburgh.
- Evers, A. 1975. *The transformational cycle in Dutch and German*. PhD thesis, University of Utrecht.
- Ferret K. 2002. *Structures syntaxiques des verbes à particule en allemand et projection aspectuelle*. PhD thesis, Université Paris 7.
- Grewendorf, G. 1991. *Aspekte der deutschen Syntax* (2<sup>nd</sup> edition). Volume 33 of 'Studien zur deutschen Grammatik', Narr, Tübingen.
- Gerdes, K. & S. Kahane. 2001. "Word order in German: A formal dependency grammar using a topological hierarchy", *Proceedings of ACL 2001*, Toulouse.
- Gibbon, D. 1998. "Intonation in German", in Di Cristo A. & Hirst D. (ed.) *Intonation systems: A survey of twenty languages*, Cambridge, Cambridge University Press.
- Haider, H. 1985. "The Case of German", in J. Toman (ed.), *Studies in German Grammar*. Studies in Generative Grammar, N° 21, Foris Publications, Dordrecht.
- Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago University Press.

- Hinrichs, E. & T. Nakazawa. 1994. "Linearizing Finite AUX in German Verbal Complexes" in Nerbonne, J., K. Netter, & C. Pollard. *German in Head-Driven Phrase Structure Grammar*, CSLI, Stanford.
- Hudson, R. A. 1980 "Constituency and Dependency", *Linguistics*, 18:3/4, 179-198; "A Second Attack on Constituency: A Reply to Dahl", *Linguistics*, 18:5/6, 489-504.
- Huybregts, R. 1976. "Overlapping dependencies in Dutch", *Utrecht Working Papers in Linguistics*, 1, 24-65.
- Kahane, S. 2002. *Grammaire d'Unification Sens-Texte : vers un modèle mathématique articulé de la langue*. Habilitation thesis, Université Paris 7.
- Kahane, S. To appear. "A propos de la position syntaxique des mots *qu-*", in P. LeGoffic (ed.), *Mophèmes en qu-*, Verbum, Paris, 44 p.
- Kathol, A. 1995. *Linearization-Based German Syntax*. PhD thesis, Ohio State University.
- Kathol, A. 2000. *Linear Syntax*. Oxford University Press.
- Lambrecht, K. 1994. *Information Structure*, Cambridge University Press, Cambridge.
- Lenerz J. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*, TBL Verlag Günter Narr, Tübingen.
- Martin Ph. 1981. "Pour une théorie de l'intonation", in M. Rossi, A. Di Cristo, D. Hirst, Ph. Martin & Y. Nishinuma (eds.). *L'intonation, de l'acoustique à la sémantique*. Klincksieck, Paris.
- Mel'cuk, I. 1965. "Porjadok slov pri avtomaticheskom sinteze russkogo teksta [Word order in automatic synthesis of Russian texts]", *Naucno-texniceskaja informacija*, 12, 36-44. [French tr.: "Ordre des mots en synthèse automatique des textes russes", 1967, *T.A. Informations*, 8:2, 65-84.]
- Mel'cuk, I. 1988. *Dependency Syntax*. State University of New York Press.
- Mel'cuk, I. 2001. *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Benjamins, Amsterdam.
- Müller, S. 1999. *Deutsche Syntax deklarativ. Head Driven Phrase Structure Grammar für das Deutsche*, 'Linguistische Arbeiten' Nr. 394, Niemeyer, Tübingen.
- Reape, M. 1994. "Domain Union and Word Order Variation in German", in J. Nerbonne *et al.* (eds.), *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, N° 46. Stanford.
- Ross J. 1967. *Constraints on Variables in Syntax*. PhD thesis, MIT [printed with the title *Infinite Syntax!* Reidel, Dordrecht, 1985].

- Skårup, P. 1975. *Les premières zones de la proposition en ancien français*. Akademisk Forlag, Copenhagen.
- Tesnière, L. 1959. *Eléments de syntaxe structurale*. Klincksieck, Paris.
- Uszkoreit Hans. 1987. *Word Order and Constituent Structure in German*. CSLI Lecture Notes 8, Stanford, CA.
- Wells, R. 1947. "Immediate constituents", *Language*, 23, 186-207.