# A New Perspective on Phrase Structure

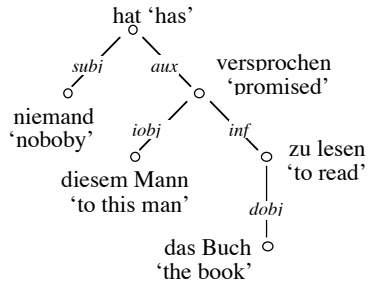
Kim Gerdes & Sylvain Kahane
Lattice, Université Paris 7, case 7003, 75251 Paris Cedex 5
kim@linguist.jussieu.fr sk@ccr.jussieu.fr

In all the phrase-structure-based approaches we know, the phrase structure encodes at the same time the subcategorization frame and the final landing points of the moved elements. We propose to separate this information and join the subcategorization frame with the communicative or information structure and to represent the linear order possibilities in a hierarchy of topological phrases. The first structure we call *syntactic structure*, the second, *topological phrase structure*.

To illustrate our point, we propose a new description of well-known data of German.[1] Our approach results in a unified description of German word order including controversial phenomena such as partial VP fronting and scrambling.

Consider the following set of sentences:

(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 diesem Mann das Buch niemand versprochen*
**g.** *?? Diesem Mann hat zu lesen niemand das Buch versprochen.*
‘ Nobody promised to this man to read the book.’




These sentences share the same meaning in the sense of assigning the same roles to the arguments. We represent this subcategorization structure by a dependency tree and show how to obtain the different corresponding word orders and phrase structures (Mel'cuk 1988). We achieve this by declaring directly the possible positions for each element depending on the topological position of its governor, i.e. its position in a hierarchy of topological domains, which are composed of fields, as in the classical topological model of Germanic (e.g. Drach, 1937, Bech 1955; Reape 1994 and Kathol 1995 for a formalization in HPSG).

Following this description of a declarative sentence, the syntactic head, the finite verb, opens the principal *domain* consisting of the following *fields*: Vorfeld, left bracket, Mittelfeld, right bracket, and Nachfeld.[2] The finite verb occupies the left bracket. A verbal argument of the main verb has two basic choices: opening a new domain or joining the domain of its governor. In the latter case, the verbal dependent takes the right bracket position of the domain of its head and it opens a limited phrase (verb cluster) with only one place for its own verbal dependent. This place is generally on its left, but can also be on its right if it is the auxiliary *werden* or *haben*.[3]

The former case is very different: the verbal argument opens an embedded domain consisting again of different fields (Mittelfeld, right bracket, and Nachfeld). It takes the right bracket of this embedded domain, and opens, just as before, a limited phrase with only a place for its verbal dependent. This embedded domain as a whole behaves just like non-verbal arguments do: both can take any place in the *major fields* (Vor-, Mittel-, and Nachfeld) of its governor’s domain or of any higher domain.[4] This

---

[1] In English or French, the subcategorization frame and the surface phrase structure are more closely related, and a merged structure may be useful. In less configurational languages like Russian, the surface order is nearly exclusively determined by the communicative structure. Finally German appears as one of the most interesting case because surface order depends strongly on both the syntactic position (e.g. finite verb in V2 or Vfinal position) and the communicative structure (e.g. content of the Vorfeld).

[2] Different restrictions apply on how many constituents can be placed in each of these fields: Vorfeld and left bracket take exactly one constituent, the right bracket at most one. For Mittel- and Nachfeld there are no restrictions on the number of constituents.

[3] Going to the left of its governor is the ordinary position of the verb and it only offers a new place to its left ($V_2V_3V_1$ isn’t possible in the right bracket). However, if it takes the (eventual) right position, the verb offers again two places, one to its left and one to its right, in order to account for simple ($?V_1V_2$, $V_1V_3V_2$, $V_1V_4V_3V_2$) and double Oberfeldumstellung or auxiliary flip ($?V_1V_2V_3$, $V_1V_2V_4V_3$) and Zwischenstellung ($?V_3V_1V_2$, $?V_4V_3V_1V_2$).

[4] The order of elements inside a domain does hardly depend on their dominance structure (but on case, pronoun vs. full noun, discourse structure, visibility of case, etc.). We propose thus to have general domain internal rules for all Mittelfelder, Nachfelder … (see, for instance, Müller, 1999: 166-175).
The creation of a new domain by bare infinitives or participles is limited to the Vorfeld.

similarity of verbal (domain) and non-verbal placement suggests a high degree of generality of our approach. Moreover, the description of Dutch is a straightforward change of some parameters.[5]
We propose in Fig. 1 to 3 different topological phrase structures for the same dependency tree. In Fig. 1 both verbal dependents join the right bracket of the main domain (V2 and verb cluster are indicated by shaded bubbles). Their non-verbal dependents end up on the same topological level and their positioning in the main fields does not depend on their position in the dependency tree (e.g. (1a) and (1b)).[6]

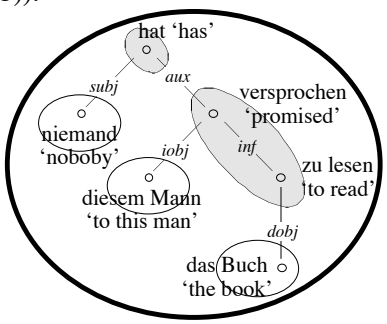



**Fig 1.** (1)a,b

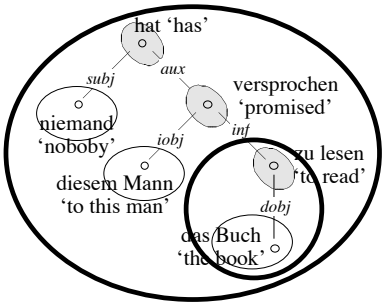



**Fig. 2.** (1)a,c,d,e

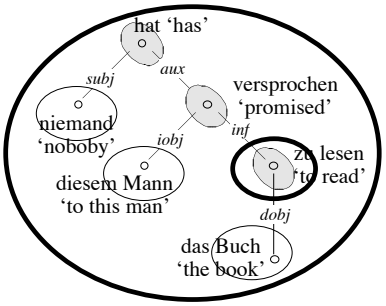



**Fig. 3**. (1)a,b,e,f,g

In Fig. 2, *zu lesen* has opened a new domain comprising *das Buch*. This embedded domain can be placed in any major field, the Vorfeld (1c), the Nachfeld (1d), or the Mittelfeld (1e). For the same phrase structure of Fig. 2, we also obtain again (1a), corresponding to a different ordering in the Mittelfeld than in (1e). This structural ambiguity for (1a) corresponds, as we believe, to a semantic ambiguity of communicative type: in Fig. 2, the fact of reading the book is marked as a semantic unit (as in *Reading the book, nobody promised that to this man*), whereas the phrase structure of Fig. 1 is neutral in this respect (*Nobody promised this man to read the book*). Moreover, the two structures correspond to different prosodies (the left border of the right bracket is clearly marked with an accent on the first syllable of the radical, and in (1a), *das Buch zu lesen* can optionally be prosodically separated just like the obligatory separation in (1e)).
Creating an embedded domain is triggered by a communicative choice. For instance, the emancipation in Fig. 3 of *das Buch* from the domain of its governor *zu lesen* must be justified by putting focus on one of the resulting elements. For the phrase structure in Fig. 3 (with four groups at the same level, *niemand*, *diesem Mann*, *das Buch*, and *zu lesen*, which, *a priori*, could be ordered freely) we therefore obtain easily only those of the possible surface orders where *zu lesen* or *das Buch* occupies the Vorfeld; word orders like (1g) oblige the reader to suppose a very specific and rare context.

We have presented an interface between syntactic structures and topological phrase structures. The syntactic structures are not linearized. They include the dependency structure and the communicative grouping, which constitutes a direct link to the semantic level of representation. The topological phrase structures on the other hand are introduced when linearizing. It allows constructing all possible word orders, and our phrases can be directly linked to prosodic units.
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 that can only accommodate one other verb, whereas a verb placed in a major domain 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 subcategorization, in accordance to language internal rules, but this information should not itself be part of phrase structure.

## References

Bech, Gunnar, *Studien über das deutsche verbum infinitum.* Kopenhagen, 1955, reprinted Niemeyer, Tübingen, 1983.

Drach, Erich, *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt, 1937.

[5] In Dutch, verb clustering excludes *te*-infinitives and the choice of 'Oberfeldumstellung' is almost always obligatory and not limited to auxiliaries. 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.

[6] 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, contrarily to English and French, is called *scrambling*. In our approach "non-scrambled" sentences do not have a simpler structure than "scrambled" sentences and do not serve to derive them (Grewendorf 1991), because we do not suppose an initial order imposed on the subcategorization structure.

Grewendorf, Günter, *Aspekte der deutschen Syntax* (2nd edition), Volume 33 of 'Studien zur deutschen Grammatik', Narr, Tübingen, 1991.

Haider, Hubert, "The Case of German", in Jindřich Toman (ed.), *Studies in German Grammar*, Studies in Generative Grammar, N° 21, Foris Publications, Dordrecht, 1985.

Kathol, Andreas, *Linearization-Based German Syntax*, PhD thesis, Ohio State University, 1995

Mel'cuk, Igor, *Dependency Syntax*, State University of New York Press, 1988.

Müller, Stefan, *Deutsche Syntax deklarativ. Head Driven Phrase Structure Grammar für das Deutsche*, 'Linguistische Arbeiten' Nr. 394, Niemeyer, Tübingen, 1999.

Reape, M., " 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, 1994.