The fields on the way to prosody
Alternatives to phrase structure based approaches to prosody

Kim Gerdes[†] **and Hi-Yon Yoo**[‡]
† Lattice, Université Parix X, France
‡ ARP, Université Paris 7, France
E-mail: kim@linguist.jussieu.fr; hi-yon.yoo@linguist.jussieu.fr

ABSTRACT

We present a model for prosodic synthesis in the framework of the Meaning-Text Theory. We discuss the usefulness of the dependency tree as the syntactic representation in comparison to the usual X-bar phrase structure, and we introduce a topological and a phonological representation as intermediate steps in the generation process. The correspondences between the different levels are formalized and implemented, and several grammars have been developed.

1. INTRODUCTION

Most studies on language generation start with a syntactic structure as the principal input of the system. As the syntactic analysis is often based on GB or minimalism, the syntactic structure is usually an X-bar phrase structure. In the setup prosody has difficulties finding its place, as it is often linked directly to other representations mainly on the semantic level.

From a practical perspective, to our knowledge, no language generation system manages to actually compute a complete S-structure from their input (that is usually some semantic representation). Thus, the work based on these structures either remains mainly theoretical or simplifies considerably the underlying theory.

From a theoretical viewpoint, it is difficult to base a prosodic structure on the S-structure, because the phrase structure encodes the whole transfer process from the logical form to the phonetic form, and its nodes can be of semantic, syntactic, or phonetic character. This is particularly problematic for languages with relatively free word order like German, Korean, or Modern Greek.

In this paper, we argue that it is important to distinguish different types of information into distinct representations, linearly ordered from meaning to sound. We propose to place the generation process of language in a Meaning-Text framework (Mel'cuk 1988) where we discuss language generation from a syntactic dependency tree, via a topological structure and a phonological structure to the phonetic output.

In section 2, we discuss the adequacy of the X-bar phrase structure. We then briefly present the Meaning-Text Theory and the different levels of representation that we consider. In section 4 and 5, we show how the linearization process constructs a topological structure from the dependency tree and how this topological structure can be transformed into a phonological structure. Finally, we illustrate this process with an example from Modern Greek in section 6.

2. X-BAR

X-bar based syntax brings with it important drawbacks for the computation of prosody:

A phrase structure obeying X-bar principles conveys two very different kinds of information: subcategorization (X, the head of the phrases X' and X'', governs syntactically the head of the non-head elements of X' and X'') and grouping (all members of X' and X" form a linear unit that is supposed to have logical, syntactic, and prosodic relevance). Of course, it has long been noticed that these different types of information do not in general coincide, a fact that has given rise to the idea of movement, i.e. elements that appear in a position where it is impossible to create a common phrase with their syntactic governor must have been "moved" to this new position from an original position that obeys to X-bar principles. The structure where the elements are in their actual linear position is called S-structure[1]. The S-structure attempts to express the linear grouping, it attempts to be easily projected to the phonological as well as to the logical form, and, of course, it still attempts to obey X-bar principles, i.e. to express subcategorization. It is clear that compromises have to be made in order to express all this information in one structure, and, likewise, it is nontrivial to extract the part of the information that is relevant for prosody. While the movement analysis can give satisfying results for the most extensively studied languages English and French, the wide gap between subcategorization and linear grouping in so-called "free word order languages" necessitates a great number of movements even in simple sentences with the implausible result that English and French seem closer than

[1] In a minimalist approach, the S-structure is merely considered as the branching point between logical and phonological form and does no longer receive a linguistic interpretation on its own.

ISBN 1-876346-48-5 © 2003 UAB

other languages to Universal Grammar.

It is well known that “free word order languages” such as Modern Greek or Russian are only free in the sense that the same argument structure can correspond to different word orders (for instance, for a sentence with a verb (V), a subject (S) and a direct object (O), the six combinations SVO, SOV, OSV, OVS, VSO, VOS are possible), but these different orderings convey in fact different communicative (or information) structures (Lambrecht 1994, Mel’cuk 2001). However, it remains unclear how this communicative information is represented and how it intervenes in the movement process.[2] Nonetheless, most of the work on the syntax-prosody link is based on GB or minimalist analyses (Selkirk 1984, Hirst & Di Cristo 1998, Rossi 1999), usually attempting to compute prosody from a (prosodically augmented) S-structure. The status of the phonological form and its relation to the different prosodic patterns remain mostly undefined.

3. THE MEANING-TEXT FRAMEWORK

We subscribe to the idea underlying science that separable information should be represented separately. This is an essential principle of the Meaning-Text Theory (MTT, Mel’cuk 1988), the theoretical framework in which we place our language generation process. In MTT, language is presented as a linear and modular process of language generation, where the different information conveyed by a sentence is separated into distinct representation levels: Mel’cuk (1988) distinguishes the syntactic, morphological and phonological levels, each of them being subdivided into a deep structure (oriented towards the meaning) and a surface structure (oriented towards the output). Starting from a semantic representation of the meaning, we construct syntactic representations (represented as syntactic dependency trees). These syntactic dependency trees are translated into morphological structures, from which phonological structures are derived that give us the language output.

Just like GB based analyses, we start from the syntactic structure and want to construct the prosodic groupings with their corresponding patterns. Note that our input structures are unordered surface syntactic dependency trees that exclusively express the syntactic subcategorization and the communicative structure of the sentence.[3] We simplify MTT’s original communicative distinction (Mel’cuk 2001) and follow Vallduvi’s idea (1992) by considering a tri-partition of the communicative structure: topic (the prominent theme or *link*), focus (the rhematic part of the sentence), and tail (non-prominent theme). All nodes of the syntactic dependency tree carry a communicative markup indicating the communicative division of the sentence.

In order to compute the prosodic structure we need an intermediate representation that contains the linear chain of words that are equipped with their prosodic characterization (such as lexical stress if any) and grouped into sequences, from which we easily compute the corresponding prosodic groupings. We follow Gerdes & Kahane 2001 and call this representation the *topological phrase structure*. The representation extends and replaces the “deep morphological representation” of classical MTT and consists of an ordered phrase structure tree, whose constituents, called domains, only express linear groupings of words and, contrary to X-bar phrase structures, do not attempt to bare syntactic information. The topological phrase structure allows us to compute the following level: the phonological structure. This is a simplification of the original MTT distinctions as we use only one more intermediate structure between topology and the sound output. This is sufficient for the computation of the prosodic contours. However, we do not exclude that a deeper analysis of morphological and phonological phenomena will necessitate a more fine-grained partition of the intermediate structures.

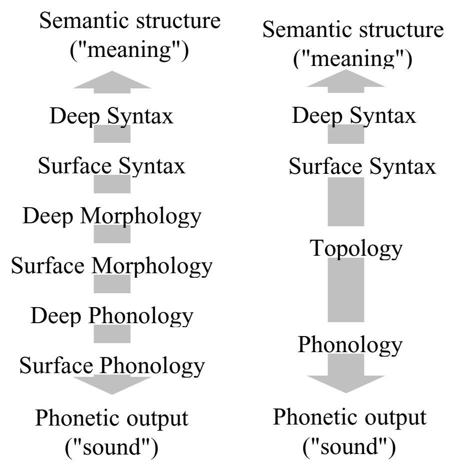

Figure 1: The levels of representation in classical Meaning-Text Theory (left) and our adaptation (right)

4. THE TOPOLOGICAL MODULE

The idea of *topology* is based on the idea of "fixed places" of a sentence called *fields* (Drach 1937, Tesnière 1959) and was initially developed for German and its highly complex word order (German is a V2 language where the finite verb has a fixed position in the sentence). The main idea is to

---

[2] For example, the fact that a given word order, say SVO, could have been triggered by different communicative structures is supposed not to show up in the S-structure, because a double analysis would be interpreted as a case of spurious ambiguity. Thus, the different possible prosodic groupings for this word order will have to be computed in the projection to the phonological form, complicating considerably this poorly formalized last step in language generation.

[3] The communicative structure is already present at the semantic level and is passed down to the surface syntactic level in the generation process.

consider a sentence as a series of fixed positions that can be occupied under certain conditions by the elements of the sentence. Gerdes & Kahane 2001 call such series of places *domains* and propose to construct a hierarchy of domains, *topological phrase structure*. Each field has its own restriction indicating the number of domains it can or has to host in order to obtain a well formed topological structure.[4]

We construct the topological phrase structure from the surface syntactic dependency tree. An element E of the dependency tree with syntactic governor G can occupy a certain field and create new domains in this field depending on E's and G's category, their communicative markup, and their topological position and depending on their syntactic relation (subject, modifier, …).

We see that both syntactic and communicative information from the dependency structure intervene on an equal basis in the linearization process. For different types of languages, the importance of the two types of information (syntactic or communicative) can vary. In the linearization rules of so-called "free word order languages", the role of the communicative information will prevail, whereas the linearization of languages like English and French is mainly driven by the syntactic relation and the category of the elements. German shows an intermediate behavior in the sense that some elements have fixed places depending on their syntactic function (e.g. the V2, the second position of the finite verb), while the placement of other elements depends on their communicative markup (e.g. the element in the *Vorfeld*, the initial position of the sentence). The resulting topological structure of English or French resembles the usual X-bar phrase structure, whereas a constituent of Greek or Russian topological structures does rarely coincide with an X-bar phrase (that expresses the syntactic relation of its elements). It is possible to establish a typology based on the nature of the topological linearization rules for different languages.

5. COMPUTATION OF THE PHONOLOGICAL STRUCTURE

The topological structure allows the computation of the corresponding phonological structure(s) without recourse to the higher structures. We describe this process in the following.

We assume that the phonological structure is a three-level tree, starting with the prosodic word as the lowest level, the prosodic group being the intermediate level, and the prosodic sentence forming the highest level. The prosodic word is the minimal prosodic unit and contains one accented word and eventually other unaccented words. It carries the lexical information, i.e. information on the segmental and suprasegmental level. The prosodic group is the intermediate level containing at least one prosodic word. We assume that the prosodic groups correspond to communicative units and that prosodic patterns are associated to the prosodic groups. Finally, the prosodic sentence contains the information on the general prosodic pattern depending on the modality of the sentence (declarative, interrogative, etc.).

The computation of the phonological structure from the topological phrase structure consists of a simplification of the topological tree. We consider two main rules in the correspondence from topology to phonology: The *prosodic word rule* that is applied on the leaves of the topological tree, allows combining unstressed words with stressed ones in order to form one prosodic word. If the word is stressed, then it founds a prosodic word. If not, it has to join one of its adjacent stressed words. Secondly, we apply the *prosodic group rules* to the domains of the topological structure. A prosodic group rule of the form (D_1, C_1, F_1, D_2, C_2, PG) allows the formation of a prosodic group PG by contraction of the domains D_1 (which has the communicative mark-up C_1) and D_2 (with the communicative mark-up C_2 and being dominated by D_1 through the field F). Domains that are not contracted by a prosodic group rule form a prosodic group on their own. With the application of these two types of rules, we obtain a hierarchical phonological structure that combines the appropriate prosodic contours with the segmental string.

The last step of language generation is the computation of the sounds of the output, i.e. we have to take into account the phonetic adjustments, and we have to compute the actual physical parameters of the prosodic contours.

6. DEPLIN

The described three-step generation process from a surface syntactic dependency tree to the sound output has been implemented in Java in a program with graphical user interface called *DepLin* (http://talana.linguist.jussieu.fr/~kim/depLin).

At present, toy grammars have been developed for French, German, Modern Greek, and Korean. In the remaining space, we present an example of the generation process starting from a Greek dependency tree.

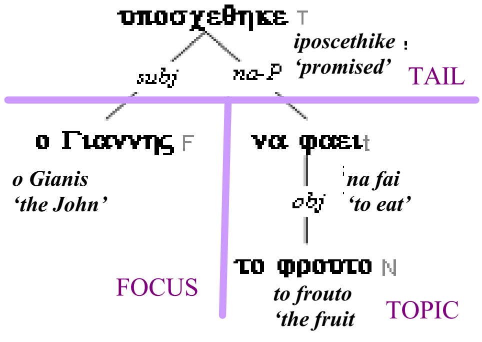

Figure 2: A Greek dependency tree with a communicative partition

[4] Note that fields are not constituents: as a consequence, two constituents that are in the same field do not form obligatorily a topological unit.

ISBN 1-876346-48-5 © 2003 UAB

The topological grammar constructs a topological phrase structure from the dependency tree in Figure 2. This grammar needs less than 10 rules for the linearization of even complex syntactic structures (verbal and phrasal complements, clitics, …). We obtain the following topological phrase structure, the only possible linearization corresponding to the given communicative markup.

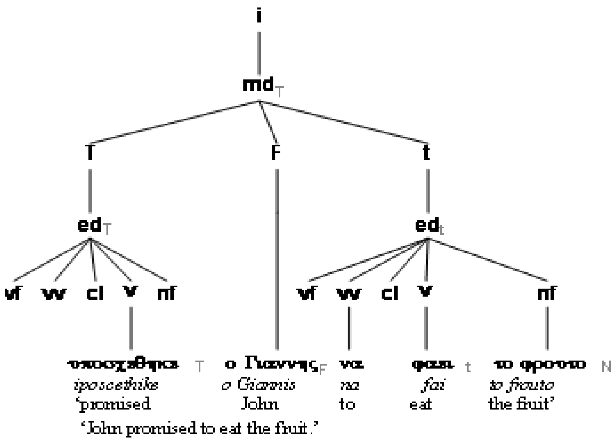

Figure 3: A topological phrase structure

From this topological tree, the phonological grammar computes the prosodic words and eventual contractions of domains that have to take place to obtain the phonological structure. We obtain the phonological structure in Figure 4. For the moment, DepLin computes only the prosodic structure, and we do not represent the internal structure of the prosodic words.

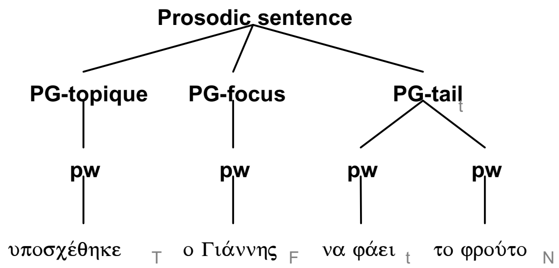

Figure 4: A simplified phonological structure

From the resulting phonological structure, we compute the sound output, for the moment by combining the appropriate prerecorded sound files for the different prosodic words. The sound output has been judged as natural by native speakers who were able to identify correctly the communicative situation of the utterance.

6. CONCLUSION

We have presented an alternative to the usual prosodic computation based on X-bar syntax. Our approach proves to be particularly suitable for free-word order languages because it naturally includes the communicative structure next to syntactic information in the linearization and grouping process of the syntactic dependency tree.

The results are very promising although the size and coverage of our grammars has not exceeded an experimental state.

Work is in progress to extend the existing grammars and to apply this process to other languages. Moreover, the translation of the leaves of the phonological tree to a phonetic string and the combination of this phonetic string with an abstract prosodic contour remains to be implemented.

REFERENCES

[1] E.Drach, *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt, 1937.

[2] K. Gerdes & S. Kahane. “Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy”, in *Proceedings of ACL 2001,* Toulouse, 2001.

[3] K. Gerdes & S. Kahane “Phrasing it differently” *Selected Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'cuk*. Benjamins, Amsterdam, Philadelphia, 2002.

[4] D.J. Hirst & A. Di Cristo (éds.), *Intonation Systems. A Survey of Twenty Languages*, Cambridge, Cambridge University Press, 1998.

[5] Lambrecht, K. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*, Cambridge Studies in Linguistics 71, Cambridge University Press

[6] I. Mel’cuk, *Dependency Syntax*. New-York, State University of New York Press, 1988.

[7] I. Mel’cuk, *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Amsterdam, Benjamins, 2001.

[8] M. Rossi, *L'Intonation : Le Système du Français: Description et modélisation***,** Ophrys, Paris, 1999.

[9] E. Selkirk, *Phonology and Syntax. The relation between sound and structure*, MIT Press, Cambridge, 1984.

[10]L. Tesnière, *Eléments de syntaxe structurale*, Kliencksieck, Paris, 1959.

[11]E. Vallduví, *The Informational Component*. New York, Garland, 1992.