Ordering Words and Placing Places

Our contribution tackles the modeling of the interaction between numerous constraints on word order. We illustrate this idea with a study of French subject and complementizer placement (standard, inverted and clitic positioning). We adopt a topological approach to word order where the positioning takes place in a structure distinct from the actual syntactic structure (the latter expressing only the functional relations, projected from the lexicon). This separation can be traced back at least to Tesnière 1959. It is showing up later in LFG and to a lesser extent in GPSG and HPSG. Authors like Milner 1989 too consider it necessary to distinguish (syntactic) position and place (in a string of words). One could even go as far as regarding the Chomskyan movement from lexical argument positions to surface positions as a syntax-topology distinction. Topological approaches are traditionally used in the description of Germanic languages (Drach 1937, Diderichsen 1953, Bech 1955, Kathol 1995, Debusmann & Duchier 2001, Gerdes & Kahane 2001) but there are few analyses of other languages (see Donohue & Sag 1999 for Walpiri and Marandin 2001 for French).

Topological analyses are often regarded as an *ad hoc* approach to word order. It is in fact possible, pushing the topological idea to an extreme, to propose a suitable template for each possible configuration without explicitly linking the different templates. Through French word order phenomena, we show that the linking of different patterns can attribute a certain explicative power to topological approaches.

We describe the syntax of a finite French verb in the following way. The finite verb constructs a *verbal domain* consisting of the following template of fields which accommodate the verb's syntactic dependents:

(1)   *comp < prefix < Vorfeld < neg-cl < me-cl < acc-cl < dat-cl < y-cl < en-cl < verb < subj-encl < adv < pred < Mittelfeld < postfix*

The *comp* field receives the possible complementizer (i.e. subordinate conjunctions and relative pronouns). The *prefix* and *postfix* fields can accommodate left and right dislocations, very common in spoken French (*Les haricots, elle n'aime pas ça, Marie*). The *Vorfeld* is open to a unique element, usually the subject, whereas the *Mittelfeld* receives any number of complements. The clitic fields (X-cl) constraints the correct placement of the clitics. The *verb* field receives the finite verb and the *pred* field can accommodate any predicative element (participle, infinitive, predicative noun or adjective) possibly separated from its head verb by an adverb.

We will now introduce a second set of constraints overlaying the constraints of (1) and organizing the verbal domain in four templates, embedded one into the other: domain, kernel, cluster, and chunk.

domain: *comp < prefix < <u>kernel</u> < postfix*
kernel: *Vorfeld < <u>cluster</u> < Mittelfeld*
cluster: *<u>chunk</u> < adv < pred*
chunk: *neg-cl < me-cl < acc-cl < dat-cl < y-cl < en-cl < <u>verb</u> < subj-encl*

The French *Vorfeld* accepts one and only one constituent (French resembles thus somehow to a V2 language). This constituent is usually the subject but other complements can occupy this place too as shows the case of the locative inversion: *Sur la place | se dresse | une cathédrale* ; *Au terrorisme | ne pouvait pas ne pas répondre | l'activisme israélien* (Le Monde, 1967).

However, when the subject is a clitic pronoun, it has to occupy the *Vorfeld*, but this field is then "aspirated" into the verbal chunk, the obligatory template for clitics (and the structure resembles more to a V1 structure). When, in this case, a locative complement is fronted it is necessarily dislocated: *Sur la place, elle se dresse*.

A comparable phenomenon of "aspiration" appears when the complementizer is also a verbal dependent: The left dislocation is possible in French completives and ungrammatical in relatives: *Je crois que, Marie, Jean ne lui donne plus d'argent* ; **l'argent que, Marie, Jean lui donne* ; *l'argent que Jean donne à Marie*. Our double system of constraints accounts perfectly for this ungrammaticality: The relative pronoun saturates the valency of the verb and must go into the kernel (as any other verbal argument would). But since it occupies the *comp* field of the domain, there can no longer be a *prefix* field, (because the *prefix* field disappears when the *comp* field is aspirated in the kernel). Moreover, the aspiration can even cause a fusion with the *Vorfeld*, the first field of the kernel, which explains the optional subject inversion in relatives (*le livre que Pierre lit* ; *le livre que lit Pierre*). This fusion is obligatory for the placement of the subject relative pronoun *qui* (**la cathédrale qui sur la place se dresse*). The same process goes even further with the interrogative pronoun *que* which is a clitic: In this case, the *comp* field is aspirated all the way into the verbal chunk causing on its way the disappearance even of the *Vorfeld* (*Que veut-il faire ?* ; **Que Pierre veut-il faire ?* ; *A qui Pierre veut-il parler ?*).

Picking up the metaphor of movement dear to Chomskyan models, we can say that in the case of aspiration, it is not the element that is moving to attain its final position, but on the contrary, it is the place that moves toward the element. A consequence of the "place movement" is the contraction of the structure and the erasure of intermediate places. We rather consider the computation of the same final (flattened out) template by means of resolution of sets of constraints than by means of movement and contraction of places described above, but the result remains the same.

Of course, chunk, cluster, and kernel also play other roles in the description. The cluster is the pivot of subject inversion in French: The inverted subject necessarily goes beyond the cluster (*le livre que veut lire Pierre* ; **le livre que veut Pierre lire*) where it scrambles freely with complements of the other verbs of the cluster (*le livre que veut donner Pierre à sa sœur*). We find the same scrambling behavior in German where the verb cluster is in final position and the complements of the verbs of the cluster mix up freely in the Mittelfeld (*dass | den Roman dieser Frau niemand | zu lesen versprochen hat*).

The chunk-cluster-kernel-domain structure is certainly neither limited to French nor to verbal structures, since we observe similar structures for nominal domains (with a nominal chunk from the determiner to the noun and a nominal cluster where even adjectives cannot be inserted: *une chemise de coton légère*; **une chemise légère de coton*).

Some references:

Debusmann R. & D. Duchier, 2001, "Topological Dependency Trees: A Constraint-Based Account of Linear Precedence", *39nd Annual Meeting of the Association for Computational Linguistics ACL*, Toulouse

Diderichsen, P. 1953, "Logische und topische Gliederung. des germanischen Satzes" In P. Diderichsen, Helhed og Struktur, 52-63. Copenhague : Gad.

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

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

Kathol A., 1995, *Linearization-Based German Syntax*. Ph.D. thesis, Stanford University

Marandin J.-M. 2001, "L'hypothèse des sites en syntaxe" in Marandin, J.-M. (ed). *Cahier Jean-Claude Milner*. Verdier, Paris.