

ON THE USEFULNESS OF MOVEMENT RULES

Petr Sgall

Prague

Abstract: If a parser is understood to have the task of achieving a syntactic representation of the input sentences, including an identification of the grammatical roles and values of the lexical occurrences, then it seems useful to distinguish between the syrface and the underlying word order (the latter corresponding to the topic-focus articulation with the scale of communicative dynamism). Certain kinds of movement rules then may be required, especially those responsible for a secondary postion of the intonation center of the sentence, the bearer of which always is the most dynamic sentence part (rheme proper). Movement rules can then also render the grammatically conditioned positions of certain kinds of words or word groups (verb, adjectival adjuncts, clitics an so on).

Keywords: parser, word order, topic, focus, communicative dynamism , systemic ordering

The goal of a parser may be understood in two different ways:

(a) to achieve a syntactic representation of the sentence in the sense of identifying the basic grammatical roles of the individual word form occurrences, so that practical applications of natural language processing such as communication with intelligent data bases, information retrieval or machine (assisted) translation become feasible, or

(b) to yield a representation of the sentence structure that would be approprite from the viewpoint of a theoretical description with a certain explanatory power, not restricted to specific kinds of applications.

ISBN: 0 08 043 438X

In case (b) the output of the analysis should contain all grammatical oppositions relevant for semantico-pragmatic interpretation. Certain kinds of movement rules then appear to be useful, as will be pointed out in the sequel on the basis of the framework of the Functional Generative Description, a synthetic view of which can be found in Sgall et al. (1986); more recently, see Hajičová (1991) and Sgall (1992); an outline of an automatic identification of Topic and Focus was presented by Hajičová et al. (1995).

As examples (1) and (2) document, the opposition of Topic and Focus (expressed by the position of the intonation center, indicated here by the capitals) is semantically relevant; in (2), even the truth conditions of the two sentences are different.

(1) (a) Yesterday it RAINED.

(a) It RAINED yesterday.

(2) (a) I study linguistics on SUNDAYS.

(b) On Sundays I study LINGUISTICS.

The role of the word order in (2) clearly differs from that in (3), although even here a movement can be found, since we can see here two options for the placement of a function word.

(3) (a) For which occasion has this been prepared?

(a) Which occasion has this been prepared for?

This example also supports the claim that function words are to be distinguished from autosemantic lexical items. Similarly as other function morphemes, function (synsemantic) words are syntactically fixed to certain lexical (autosemantic) words - prepositions to nouns, conjunctions and auxiliaries to verbs. It is not adequate (since not economical) to represent them by nodes of the tree. As far as they are immediately semantically relevant, they can be indicated by indices of complex node labels (morphological grammatemes - values of tense, aspect, modalities, number, definiteness). As far as they are relevant just for grammatical agreement, they concern only the relationship between the syntactic representation (which has the shape of a tree, or, taking coordination into account, of a more complex network, see Petkevič 1987) and the morphemic one (which is a string of symbols). The latter case is that of gender, number and case with adjectives in many languages, as well as of person, number with verbs, the (surface) symbols for which are present only in the morphemic string.

Examples such as (1) or (2) support the view that the prototypical means of expression of Topic and Focus can be seen in an interplay of surface word order and sentence prosody (if specific morphemic and syntactic means are understood as more or less peripheral). It is therefore useful to distinguish the underlying word order from the surface one. The former can be understood as correlated with the **communicative dynamism** (CD), studied in the context of the topic-focus articulation. CD differs from surface word order first of all in the fact that the bearer of the intonation center (which, on the surface, may occupy different positions) is the most dynamic part of the sentence.

ISBN: 0 08 043 438X

In languages exhibiting a higher degree of "free" word order, the prototypical position of the intonation center at the end of the sentence is even more frequent than in English, which is made possible e.g. by the subject being able to follow after the verb, cf. the Czech translations of (4) and of (5), the latter being more or less equivalent to (2b):

(1) Then two NEIGHBORS came in.
Cz. Pak vešli dva SOUSEDÉ.
lit. Then came-in two neighbors.

(2) I study LINGUISTICS on Sundays.
Cz. V nedìli studuju LINGVISTIKU.
lit. On Sunday I-study linguistics.

It should be noted that the "free" word order is free in not being constrained by shallow grammatical rules, but it is determined by CD.

The opposition of Topic and Focus is relevant for quantifier scopes, as illustrated by the following example (we use a German sentence, since the degree of "free" word order in German is high enough for such a sentence as (6b) not to require passivization, as is the case in English):

(6) (a) Viele Leute lesen wenige Bücher.
lit. Many people read few books.

(b) Wenige Bücher lesen viele Leute.
lit. Few books read many people.

In this and the further examples we do not indicate the bearer of the intonation center if this is the rightmost sentence part.

Even within topic the degrees of CD are similarly relevant:

(7) (a) Daß viele Leute wenige Bücher lesen, ist typisch.
lit. That many people few books read is typical.

(b) Daß wenige Bücher viele Leute lesen, ist typisch.
lit. That few books many people read is typical.

(8) (a) It is JOHN who spoke to few girls about many problems.

(b) It is JOHN who spoke about many problems to few girls.

Moreover, this underlying order can be viewed as reflecting an important constraint: in the focus part of every sentence (in English, French, Czech, Russian, etc.) it corresponds to a certain ordering of valency slots defined by the grammar and present in the valency frames (grids) of lexical entries. This is what in the Functional Generative Description is called **systemic ordering** (SO; see Sgall et al. 1994 and the publications quoted there; in the research on 'scrambling', as carried out in the context of different schools, the results of this inquiry have not yet found a full reflection), and what can be illustrated as follows:

ISBN: 0 08 043 438X

(9) (a) They moved from a village to an industrial center.

(b) They moved to an industrial center from a village.

In (9a) the two Directionals both are in its focus on some of its readings, and on other readings the *from*-group belongs to the Focus, but the *to*-group to the Topic. On the other hand, in (9b) the *to*-Directional can only be included in the Topic, as has been checked by several series of experiments with tens of speakers (up to now these experiments could be only limited, but it is an urgent task to continue this research for individual languages at a much larger scale).

Similarly the presence and relevance of SO can be specified for other pairs:

(10) (a) They went by car to the river.

(b) They went to the river by car.

The secondary character of (10b), with which the *to*-Directional belongs to Topic on all readings, shows that the complementation (adverbial, adjunct) of Manner precedes the *to*-Directional under SO.

(11) (a) Jim dug a ditch with a hoe.

(b) Jim dug a DITCH with a hoe.

Since in (11b) the complementation of Means can only belong to Topic, under SO Objective precedes Means. The switch between (a) and (b) concerns intonation here, rather than the surface word order position, since the degree of "free" word order in English is not high enough to allow for the Objective to follow Means, although, as we have seen with the preceding examples, the English word order is "free" enough to make a switch of two adverbials possible in certain cases; other "free" word order options characterize the next two examples.

(12) (a) *Ron cannot sleep quietly in a hotel.*

(b) *In a hotel Ron cannot sleep quietly.*

Under SO Manner precedes Locative, which follows from the fact that in (12b) Locative, placed to the left, can only be included in Topic.

(13) (a) Dutch companies published many books on linguistics.

(b) Many books on linguistics were published by Dutch companies.

With passivization the Objective, placed - as subject - to the left, belongs to Topic (with normal intonation, with which the intonation center is placed on the rightmost sentence part), and thus Actor precedes Objective under SO.

The **surface word order** can be derived from the underlying one by movement rules responsible for those cases in which a secondary placement of the intonation center

ISBN: 0 08 043 438X

corresponds to a marked position of focus (not in the rightmost position), as well as for grammatically determined positions of word forms belonging to specific groups, such as:

(a) verbs (after the subject, etc., in English, in the "second position" in German, and so on),

(b) adjectives (before their head nouns in English, German, Czech, etc., although prototypically after them in French),

(c) clitics (in Wackernagel's position in some languages, after the finite verb in others), and so on.

Also the positions of function words may be specified by movement rules concerning the difference between CD and the surface word order. Usually, function words are placed at the beginning of their word groups: prepositions and articles before nouns (and adjectives), conjunctions before verb groups, grading particles (e.g. English *more, most*) before adjectives or adverbs. In other cases (as with auxiliary verbs) the surface positions must be determined in more complex ways.

During parsing, the assignment of underlying (i.e. CD) positions to function words is relatively simple if - as is done in the Functional Generative Description - their underlying counterparts are handled as parts (indices) of complex node labels of the corresponding autosemantic units. To cope with the other differences between the surface word order and the scale of CD, we can use numerical indices for the underlying positions (as S. Starosta, p.c., proposes). This may be highly useful for a parser that should be as economical as possible. However, to obtain a full - and perspicuous - underlying representation, it is perhaps more adequate to move the word occurrences to their CD positions, so that e.g. *rained* in (1b) or *neighbors* in (4), *linguistics* in (5), etc., is moved to the rightmost position, as the most dynamic element of the sentence.

The dependency trees representing the underlying sentence structures can then be linearized, if the order of the lexical occurrences is that of CD and the main verb has a superscript t whenever it belongs to the topic, and if every dependent sentence part is enclosed in a pair of parentheses; the kind of dependency (the valency slot, i.e. the type of argument or adjunct) can be marked by an index on that parenthesis that is oriented towards the head word. (As for indices corresponding to morphological values such as those of number, tense, etc., see Sgall et al. 1986; here they are highly simplified, as is the indication of the lexical units.) Thus we obtain the following linearized trees for some of our examples:

(1') (a-b) $(\text{yesterday})_{\text{Temp}}$ rain.Pret.Declar

(2') (a) $(\text{I})_{\text{Actor}}$ study.Pres.Declar $(_{\text{Obj}}$ linguistics) $(_{\text{Temp}}$ Sunday.Plur)

(b) $(\text{Sunday.Plur})_{\text{Temp}}$ $(\text{I})_{\text{Actor}}$ study.Pres.Declar $(_{\text{Obj}}$ linguistics)

(8') (a_1) (girl.Plur $(_{\text{Restr}}$ few$))_{\text{Addr}}$ (problem.Plur $(_{\text{Restr}}$ many$))_{\text{Obj}}$ speak$^{\text{t}}$.Pret.Declar $(_{\text{Actor}}$ John)

ISBN: 0 08 043 438X

(a_2) (girl.Plur ($_{Restr}$ few))$_{Addr}$ speak^t.Pret.Declar (problem.Plur ($_{Restr}$ many))$_{Obj}$ ($_{Actor}$ John)

(b) (problem.Plur ($_{Restr}$ many))$_{Obj}$ speak^t.Pret.Declar (girl.Plur ($_{Restr}$ few))$_{Addr}$ ($_{Actor}$ John)

(10') (a) (they)$_{Actor}$ go.Pret.Declar ($_{Means}$ car) ($_{Dir\text{-}to}$ river.Def)

(b) (they)$_{Actor}$ go.Pret.Declar ($_{Dir\text{-}to}$ river.Def) ($_{Means}$ car)

(11') (a) (Jim)$_{Actor}$ dig.Pret.Declar ($_{Obj}$ ditch) ($_{Means}$ hoe)

(b) (Jim)$_{Actor}$ ($_{Means}$ hoe) dig^t.Pret.Declar ($_{Obj}$ ditch)

One more example may be useful to illustrate different kinds of movement.

(14) Jim returned for his old umbrella since it was raining.

(14') (Jim)$_{Actor}$ ((he)$_{Appurt}$ umbrella ($_{Restr}$ old))$_{Purp}$ return ($_{Cause}$ rain)

(e.g. as an answer to: Did John have his umbrella with him?)

The abbreviations are to be read as Appurt(enance - broader than "Possessive"), Restr(ictive general adjunct), Purp(ose), Obj(ective).

The movements in this example (all of which are oriented from the right to the left) concern:

(i) the position of *return* - in the surface word order of English the verb regularly precedes most adverbials;

(ii) the movement of *old*, which, as an adjectival adjunct precedes its head noun;

(iii) the position of the function words *for, since, it, was*, which move to their typical positions.

Movement rules can also help in the description of the seemingly non-projective constructions (in other words, those in which the condition of adjacency is not met), if they can be handled as projective on the level of (underlying) syntax, although occupying other positions in the morphemic representation. Since the latter has the shape of a string, rather than of a tree or of another network, the condition of projectivity (adjacency) is not relevant for them. Thus, in the following examples, we suppose that the dependency trees (syntactic representations) can be seen as projective, only the corresponding morphemic strings displaying the seemingly non-projective word order:

(15) A man CAME IN, who wore a straw hat.

(15') (man$_{Actor}$ ((who)$_{Actor}$ wear.Pret.Declar ($_{Obj}$ (straw)$_{Restr}$ hat)) come-in.Pret.Declar

ISBN: 0 08 043 438X

(16) Him they intended to send to FRANCE.

(16') $(\text{they})_{\text{Actor}}$ intend.Pret.Declar $(_{\text{Obj}}\ (\text{he}^{c})_{\text{Obj}}$ send $(_{\text{Dir-to}}$ France))

(17) Amherst is a SMALLER town than Boston.

(17') (Amherst) is (town ((Boston) smaller)

The movement rules illustrated here can be briefly characterized as the 'heavy constitutent' movement in (15), a movement of a contrastive part of the topic to the left (or 'topicalization') in (16), crossing the 'higher' verb, and a local movement conditioned by the given grammatical construction (with the Comparative) in (17). The basic hypothesis is that all movements rules can be specified - in the two directions, for synthesis or generation as well as for parsing - by strict grammatical conditions, although cases of ambiguity or synonymy (i.e. of optional movements) are not excluded.

REFERENCES

Hajičová E. (1991) "Free" word order described without unnecessary complexity. *Theoretical Linguistics* **17**:99-1O6.

Hajičová E., P. Sgall and H. Skoumalová (1994) An automatic procedure for topic-focus identification. *Computational Linguistics* **21**:81-94.

Petkevič Vladimír (1987) A new dependency based specification of underlying representations of sentences. *Theoretical Linguistics* **14**:143-172.

Sgall P. (1992) Underlying structure of sentences and its relations to semantics. In: Wiener Slawistischer Almanach, Sonderband 33. Ed. by Tilmann Reuther, 273-282.

Sgall P., E. Hajičová and J. Panevová (1986) *The meaning of the sentence in its semantic and pragmatic aspects*, ed. by J. L. Mey. Dordrecht:Reidel - Prague:Academia.

Sgall P., O. Pfeiffer, W. U. Dressler and M. Pùček (1994) Experimental research on Systemic Ordering. *Theoretical Linguistics* **21**, 197-239.

ISBN: 0 08 043 438X