I. szám, 1985. november

LÁSZLÓ KÁLMÁN - GÁBOR PRÓSZÉKY

FMR GRAMMAR*

Az MTA Nyelvtudományi Intézete
1250 Budapest, Szentháromság utca 2.

Az MTA Zenetudományi Intézete
1014 Budapest, Táncsics Mihály utca 7.

FM in FMR stands for the Hungarian Fickó és Mackó, two figures in a Hungarian children's review whose names never appear in writing, drawings are used instead, even within sentences. Since Hungarian children are not supposed to have any lexical information in the linguistic sense about these drawings, the question arises how they are able to interpret the sentences containing them, if the grammar they use makes crucial use of lexical categories. The paper claims that such phenomena are not at all exceptional in the use of natural languages. The authors outline a type of non-constructive grammar where lexical information is taken into account in semantic processing only, the syntactic rules being independent of any lexical knowledge.

* We are indebted to Zoltán Bánréti and András Kornai for their comments on earlier versions of this paper.

0. Introduction

The present paper gives an overview of a possible model of linguistic knowledge, one that differs basically from the model generally accepted today, brought into fashion by generative grammar. It is impossible to prove purely formally either that the concept of language as a stringset is unsatisfactory, or that the concept that we propose is the best, or indeed the only, alternative. All that we propose is to outline this alternative and present some evidence for its superiority over the stringset model.

The core of our conception* is that grammar, devoid of both intelligence and knowledge of the external world, is not in a position itself to decide whether a given utterance makes communicative sense. Utterances are not simply the output of a language assembly line functioning in the output of a language assembly line functioning in the interlocutors' minds, with grammar for its rules of assembly, but rather manifestations of their use of the entire assembly system.

Can an unintelligent language user decide whether a given utterance is 'acceptable' or not? Generative linguistics answers this question in the positive. In our opinion, however, a serious linguistic concept should not rely on assumptions of this kind. That is, decisions concerning the 'acceptability' of an utterance should not be assigned to grammatical competence alone; consequently, neither should a grammar be expected to provide sufficient criteria for such decisions.

1. Theoretical Background

1.1. Language: Another Aspect

As is well-known, formal grammars identify language with the set of utterances that may potentially occur in the course of its use. Accordingly, their prime concern is to make an exhaustive list of these with the aid of (1) a finite vocabulary (or alphabet) V and (2) rules that select acceptable utterances from the set of all finite strings over the vocabulary (L is contained by V^*).

* For earlier papers on FMR grammar see Kálmán (1983) and Prószéky (1984).

In our opinion, however, grammar should be seen not as a mechanism producing all possible utterances, but as the knowledge by which interlocutors recognize linguistic structures. Since all elements of an utterance are not of equal relevance for structure marking, a grammar can clearly provide a satisfactory structural characterization of all possible sentences without providing an algorithm for the exhaustive listing of all of those. This is why the grammar we propose is n o n - c o n s t r u c t i v e.*

We posit that languages differ to the extent that the non-constructive grammars that can be assigned to them differ. Other aspects of linguistic knowledge, such as the knowledge of the vocabulary or the strategies of utterance analysis and synthesis, will be considered secondary here, even though they, too, vary as a rule from one language to the next. The strategies are evidently based on features of the non-constructive rule system, some but not necessarily all of which are language specific, e.g. the importance of word order, the type of morphemes and suprasegmental signs encoding syntactic structures and boundaries, etc. in the given language. The vocabulary, however, may be regarded as independent of these rules, although intelligent speakers do certainly adapt strategies to vocabulary in most of the cases.

1.2. FMR Grammar

The non-constructive grammar we propose, called FMR grammar, cannot produce and/or analyse utterances in itself - its function is limited to storing information concerning the relevant features of possible utterances in terms of both the generating and the parsing strategies. Accordingly, its rules cannot include concepts upon which no unambiguous identification strategies can be based, such as empty categories, transformation and deletion rules, etc.

The FMR rule system has a close structural resemblance to a system of rewrite rules. However, since syntactic and semantic rules are coupled according to the "rule-to-rule hypothesis" of Bach (1976:2), phrase markers are inactive by-products of the derivation; they are not "interpreted" in any way. This allows a very broad use of surface patterns (see Kálmán and

* For the term non-constructive grammar see Langendoen and Postal (1984) who, however, base their contention that language is not a stringset on the alleged existence of transfinite sentences.

Kornai (in prep.)).

Syntactic rules have the form

$$L : R_1 \; R_2 \; \ldots,$$

that is, a string is L only if its part can be considered R_1, its second part can be considered R_2, and so on. Any R is an element of exactly one of the following sets: (1) rule names (of which L itself is one); (2) preterminal symbols; (3) variables. If R is a rule name, then there exists a rule of the form R : ... If R is a preterminal symbol, then the entities with the features that it specifies are exclusively lexical items. Otherwise, R is a variable ranging over arbitrary (phonologically well-formed) terminal symbols. The elements of sets (1) and (2) are complex feature structures (cf. Gazdar et al. (1984)). The form of the semantic rules is immaterial for the present discussion (for details, cf. Kálmán (in prep.)). These can refer to symbols R_1, R_2, ... and, importantly, they can assign one function to each variable of the rule. Before substituting a variable by elements of the actual lexicon, it should be investigated first how the lexical item in question can be assigned the function specified in the semantic rules; the substitution often requires the augmentation of the actual lexicon.

Generative grammars consider a sentence X whose elements figure in the vocabulary but which cannot be generated by the rule system (X is from V^*, X is not in L), and a sentence Y that includes elements absent from the lexicon (X is not in V^*) equally ungrammatical. FMR grammar necessarily does not consider a string of the latter type ungrammatical: it rules it out only if the strategies cannot (or do not want to) assign any structure to it on the basis of the non-constructive rule system. The ignorance of a rule certainly impairs linguistic competence much more than the ignorance of a lexical item.

Since variables range over arbitrary elements, FMR grammar is, in principle, much more tolerant than either a generative rule system or even the average speaker. We suppose, however, that the strategies driving the FMR rule-system are reluctant to stretch too far the possibilities inherent in the non-constructivity of the rules, e.g. to assume that the speaker is using too many unknown elements or to use too many of them. The ignorance of a lexical element may entail refusal on this level only; this, however,

however, has absolutely no bearing on grammatical knowledge.

2. Practical Considerations

2.1. Stringset or Capacity?

It is well-known that we can recognize and can make our audience recognize the structure of nonsense sentences, as in the case of jabber, an irrefutable instance - albeit a rather extreme one - of language usage, a thesis clearly proven by such lines of poetry as

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mone raths outgrabe.

2.2. Uneven Relevance of Elements

The jabber phenomenon confirms our thesis that the elements of a construction are not equally relevant for structure identification. Let us replace the structurally irrelevant elements in the above example by asterisks:

'Twas *, and the * *s
Did * and * in the *:
All * were the *s,
And the * *s *.

Any English-speaking person can formulate the following hypotheses about the structure of the above sentences:*

'Twas $\left\{\begin{matrix} \text{N} \\ \text{Adj} \end{matrix}\right\}$, and the Adj Ns
Did V and V in the N:
All Adj were the Ns,
And the $\left\{\begin{matrix} \text{N} & \text{Vs}\left\{\begin{matrix} \text{N} \\ \text{Adv} \end{matrix}\right\} \\ \text{Adj} & \text{Ns} \quad \text{V} \end{matrix}\right\}$.

* Our thesis is unaffected by the fact that jabber often allows several alternative interpretations: the same holds for correct "real" utterances excepta that, there, dilemmas are unlikely to arise, thanks to the familiar meanings of lexical elements, provided the context too is sufficiently familiar.

The context and various other clues will prompt the reader to decide in favour of the following interpretation:

'Twas Adj, and the Adj Ns
Did V and V in the N:
All Adj were the Ns,
And the Adj Ns V.

Here again, the argument is not in the least influenced by the fact that jabber is, after all, a peculiar linguistic construct. The non-uniform structural importance of elements is demonstrated by everyday linguistic phenomena also: In Hungarian, the elements appearing in the context Láttad ...-t? 'see (Past 2nd Sg definite-object) ...(ACC)' can have no structure-marking function at all (irrespective of realization) and do not even determine their own function: they will unconditionally be interpreted as proper names: 'Did you see X?' It would be redundant to trace back the function of the inserted element to both syntactic structure and lexical category.

2.3. A Fixed Lexicon?

Any member of a language community has a finite vocabulary at any given time. Although these vocabularies change all the time, the changes that they undergo do not in the least influence language knowledge. Although a few years ago there existed no string gorbatchev in average Hungarian awareness, the first utterance which contained it was immediately acceptable to any Hungarian speaker. It is no use distinguishing between two languages simply because $V_2 = V_1 \cup \{\text{gorbatchev}\}$. We intend no treatment of the lexicon flexible enough to allow the introduction of certain new elements from time to time, rather, we claim that the lexicon is not a prime defining feature of a language.

3. Illustration

Let us now review a few FMR rules for the Hungarian. The mechanisms governing feature percolation and semantic interpretation will both be encoded in the complex symbols of the rules. The rules describe the outlines of certain types of Hungarian NP. The terms of the rules have a tree-like structure. The roots of the trees are rule names and preterminals; the tree nodes are labelled by features and linearized by bracketing. The value of each feature within a complex symbol is the subtree that it dominates. As to the feature-marking conventions, let it suffice to say that (1) it is not necessary for a complex symbol to cover all the branches of a tree; (2) it is not necessary to indicate the value +; (3) the feature - can have no value (see Kornai (1984)).

As mentioned above, the rules comprise the semantic rules implicitly: a complex symbol may include the feature SEM, whose value describes the semantic representation. All features with the exception of the SEM subtree are syntactic.

(1) (N (BAR 3) X (SEM Y) (CASE Z)) :
(N (BAR 2) X (SEM Y)) + (CASE-ENDING Z).

In modern syntax, this rule would be a lexical one (and could actually be recorded in the lexicon; yet we assume here that the lexicon contains no rules). CASE-ENDING is a preterminal symbol, with lexical representations such as

<u>0</u> (CASE-ENDING SUBJECT)
<u>at</u>/<u>et</u>/<u>ot</u>/<u>öt</u>/<u>t</u> (CASE-ENDING OBJECT)
<u>nak</u>/<u>nek</u> (CASE-ENDING DATIVE)
etc.

(2) (N (BAR 2) X (DEFINITE Y)
(SEM (APPLY Z V))) :
(DETERMINER (DEFINITE Y) (SEM Z) + (N (BAR 1) X (SEM V)).

DETERMINER is a preterminal symbol:

<u>a</u>/<u>az</u> (DETERMINER DEFINITE (SEM (PRESUPPOSE X)))
<u>minden</u> (DETERMINER (DEFINITE -) (SEM (UNIVERSAL X)))
etc.

(3) (N (BAR 2) (PLURAL -) (DEFINITE -)
(SEM (NUMBER X Y))) :
(NUMERAL X) + (N (BAR 0) (SEM Y)).

NUMERAL is preterminal:

egy (NUMERAL 1)
két (NUMERAL 2)
etc.

(4) (N (BAR 2) (PLURAL -) DEFINITE
(SEM (PROPERNAME alpha))) : alpha.

where alpha is a variable whose actual role is assigned to it by PROPERNAME. A definition of PROPERNAME may be something like (PRESUPPOSE (HIS-NAME-IS X)). If John, e.g., is an unknown element in a sentence to be parsed, its translation will be equivalent to that of the one called John, and a lexical item of the form

John (N (BAR 2) (PLURAL -) DEFINITE (SEM (PROPERNAME JOHN)))

can be created. Such representations, however, are idle, because a correct syntactic analysis is feasible. In other cases idiosyncratic anomalies make the lexical treatment helpful:

a Gázművek (N (BAR 2) (PLURAL -) DEFINITE (SEM (PRESUPPOSE
GAS COMPANY)))

The word could be derived by morphological rules only if it were (PLURAL).

(5) (N (BAR 1) (PLURAL -) (SEM (NUMBER X Y))):
(NUMERAL X) + (N (BAR 0) (SEM Y)).

(6) (N (BAR 1) PLURAL (SEM (SEVERAL X))) :
(N (BAR 0) (SEM X)) + (PLURAL-ENDING).

(6) is another rule which would have to be in the lexicon if we used a lexical phonological treatment. PLURAL-ENDING is a preterminal symbol:

ak/ek/ok/ök/k

(7) (N (BAR 0) (SEM (NOUN alpha))) : alpha.

Derivational rules can be built into the syntactic rule system in much the same way as inflect ional ones. For an example, consider (8):

(8) (BAR 0) (SEM (APPLY X Y))) :
(VERB (SEM Y)) + (VERB-TO-NOUN-SUFFIX X).

(9) (N BAR 1) X) : (N (BAR Ø) X)

VERB-TO-NOUN-SUFFIX is a preterminal symbol:

ás/és (VERB-TO-NOUN-SUFFIX (NOMEN-ACTIONIS X))

In addition to the above, in Hungarian (N (BAR 0)) can, unlike verbs, be rewritten by a variable. This interesting restriction is not universal. English is one counterexample; such phenomena are impossible to be expressed in a conventional generative framework.

4. Concluding Remarks

Formally speaking, FMR gramm ar is an extension of traditional (non-transformational) generative grammar: the main difference is that in generative grammars the range of the variables is fixed. The generative power of FMR grammar, however, is infinitely greater than that of a context-free phrase structure grammar. This is why we emphasize that the FMR rule system is not considered an automaton itself. The language it generates together with the extra-linguistic knowledge (which includes the knowledge of vocabulary, parsing and generating strategies, etc.) is likely to be regular (see Kálmán and Kornai (in. prep.)).

The extra-linguistic knowledge--FMR grammar interaction is crucial in other respects also. The concept outlined in the present paper offers a natural procedure for augmenting the actual lexicon (which is considered extra-linguistic) whenever an unknown element appears in a construction to be parsed.

The procedure for enriching the FMR rule system may be less self-evident. An enrichment can be implemented in the following manner: a new rule can be created whenever a rule name, expanded as a variable, is realized by a string of lexically well-established items in an actual construction to be parsed.

By way of an example, suppose that rules (1-8) hold, and we are to parse the construction a ravaszdi róka 'the cunning fox'. The only applicable rule is (2), and then (9). However, provided ravaszdi (ADJ (SEM (CUNNING)) and róka (N (BAR §) (SEM FOX)) are well established lexical items, it is reasonable to enlarge the rule system by

(10) (N (BAR 0) (SEM (ATTRIBUTE X Y))) :
(ADJ (SEM X)) + (N (BAR 0) (SEM Y)).

In practice, this move requires for the parser to know in advance what ravaszdi róka means (if it does not, other analyses are available; it is assumed that natural language speakers fulfil this requirement whenever they actually enlarge their rule system.

Irodalom

Bach, E. (1976), An Extension of Classical Transformational Grammar. Mimeo, University of Massachusetts at Amherst

Gazdar, G. - Klein, E. - Pullum, G. - Sag, I. (1984), Generalized Phrase Structure Grammar. Manuscript.

Kálmán, L. (1983), Un ensemble de propositions pour la science du langage. Kodikas/Code 3-4: 237-244.

Kálmán, L. (in prep.), Corrective sentences in file change semantics.

Kálmán, L. - Kornai, A. (in prep.), Pattern watching: The context-free treatment of 'context-sensitive' phenomena.

Kornai, A. (1984), Lexical features and X-bar categories. To appear in: Acta Linguistica Academiae Scientiarum Hungaricae 35.

Langendoen, T. - Postal, P. (1984), The Vastness of Natural Languages. London, Blackwell

Prószéky, G. (1984), Review on Langendoen and Postal (1984). To appear in: Studies in Language.