Relational Phrase Structure Grammar Applied to Mohawk Constructions*

IRENA BELLERT

A formal system, which is interpreted as a grammar of a language and called "relational phrase structure grammar", is applied to certain constructions of Mohawk. The Mohawk constructions contain cross-dependencies because of which Mohawk cannot be generated by a phrase structure grammar, as defined in the literature.

Recently much concern has been given to the problem of the generative adequacy of different types of grammars, [1], [2], [3], [4], [5], [6] and [7]. There are certain types of dependencies in languages which constitute a crucial test for grammars in this respect; languages with certain types of dependencies cannot be generated by certain grammars, while they can be generated by others. For instance, self-embedding properties constitute such a test, as it has been proven by Chomsky (see Theorem 33 in [4]) that self-embedding is a crucial property distinguishing finite state languages from phrase structure languages.

Another property which is essential in distinguishing context-free from context-sensitive languages is that of unbounded cross dependencies. One of the languages possessing this property is, for instance, Chomsky's L_3, which contains all sentences consisting of a string followed by that indentical string, and only these. ([5] p. 285). Chomsky has shown that there exists no context-free grammar for that language, and has presented a context-sensitive grammar for it ([4] p. 367, Theorem 15). However, Chomsky's context-sensitive rules for that grammar do not satisfy the condition that permutations be excluded in derivations, which condition has to be satisfied for a meaningful phrase structure representation of natural languages ([7] p. 144).

The problem of cross dependencies has aroused linguistic interest by the evidence

* This paper was written when the author was working at the Department of Linguistics of the University of Pennsylvania on the project sponsored by N.S.F.

provided by Postal in connection with Mohawk [7]. Postal has demonstrated that Mohawk lies outside the bounds of context-free description, by showing that it contains an infinite set of sentences with the formal properties of the language L_3. Let us quote Postal's description of certain structure in Mohawk, which display the cross dependencies in question.

"A simple Mohawk sentence consists of a subject noun, a verb, and an object noun proceded by the particle *ne-* in that order.* A noun consists of a noun prefix, a noun stem, and a noun suffix, which is often phonologically null. A verb consists minimally of a pronominal prefix, a base and a suffix. Hence there are sentences like:

kaksa?a kanuhwe?s ne-kanuhsa? the girl likes the house
1| 2 3| 4 |5 6 | 7 |8

1. noun prefix	5. suffix — serial aspect
2. noun stem *child*	6. noun prefix
3. pronominal prefix	7. noun stem *house*
4. base — verb stem *to like*	8. noun suffix

It should be noted that the Base constituent of the verb is here simply a single verb stem. However, the Base consistituent may also consist of an incorporation marker plus a verb stem. That is, one of the rules of Mohawk grammar is:

$$\text{Base} \rightarrow \text{(inc) Verb Stem.}$$

And there is an interesting process which incorporates the noun stem of the object of a sencence (like above) into the verb by substituting the noun stem for the incorporation marker. In these cases the Base consituent then has the structure:

$$\text{Noun Stem Verb Stem}.$$

Thus, there are sentences like:

kaksa?a kanuhsnuhwe?s the girl likes the house

The crucial fact about incorporation from the present point of view is that under certain conditions (the condition is that a Modifier precede the noun whose stem is to be doubled), incorporation occurs in such a way that the external noun of the object is also present. One therefore finds sentences like:

kaksa?a kanuhwe?s kikʌ kanuhsa? the girl likes this house

In such sentences there is a strict dependency ... between the incorporated noun

* For a more complete description of the relevant aspects of Mohawk sentence structure Postal refers the reader to his doctoral dissertation "Some Syntactic Rules in Mohawk" (Yale University Doctoral Dissertation, 1962).

stem and the noun stem of the external object noun ... Sentences can contain both incorporated noun stems and external noun objects only if the noun stem of the object is the same as the incorporated noun stem.

Almost any Base constituent can be nominalized by the addition of a nominalizing morpheme [*hsra/tsra*] to derive and abstract neuter noun stem. ...This nominalization can occur as well with Base constituents that contain incorporated noun stems ... Now complex noun stems produced by this nominalization are themselves capable of renominalization, of reincorporation and so on. Thus, the process of abstract noun stem formation is recursive, and there is no bound on the length of noun stems, and therefore no limit to the number of nouns".*

Thus Postal has proved in his paper that Mohawk cannot be enumerated by a context-free phrase structure grammar, and he has given transformational rules for the Mohawk constructions in question. Moreover, it was pointed out in the paper that a context-sensitive description constructed according to Chomsky's device (which violates the condition that permutation be excluded) "would reduce the phrase structure description of Mohawk sentences to absurdity" and "at least $6n^2$ rules would be required to enumerate the construction, where n is the vocabulary involved in the equivalence" ([7] p. 150).

We shall now present a relational phrase structure grammar, which will enumerate the above Mohawk construction, and assign a phrase structure marker to each sentence.

We shall first present the definition of a relational phrase structure grammar,** and then we shall appy this grammar to Mohawk.

A relational grammar is the system

$$\langle \hat{Z}, C^i, \hat{G}^i \rangle ,$$

where $i = 1, \ldots, r$, and

$\hat{Z}$ is a simple*** phrase structure grammar $\langle V, P, T, S \rangle$, where V is the entire vocabulary, P is the set of production rules, T is the terminal vocabulary and S is the initial symbol;

C^i is an element of $V - T$ (that is, non-terminal vocabulary);

$\hat{G}^i_1 = \langle\langle G^i_1, G^i_2, \ldots, G^i_k\rangle, R^i\rangle$ is a system of coupled simple phrase structure grammars, where $\langle G^i_1, G^i_2, \ldots, G^i_k\rangle$ is a k-tuple of grammar, and R^i is a relation of coupling included in $P^i_1 \times P^i_2 \times \ldots \times P^i_k$ (P^i_j being the set of production rules of G^i_j), that is, R^i is a given set of k-tuples $\langle p_1, p_2, \ldots, p_k\rangle$ such that p_j is a production rule belonging to P^i_j (for $j = 1, \ldots, k$).

* This passage was quoted from Postal, [7] p. 147–8.

** A relational phrase structure grammar has been described in my doctoral dissertation (Warsaw University, 1964), and in [8].

*** The term "simple phrase structure grammar" is used here as an equivalent to context-free phrase structure grammar.

k-tuples of sentences generated by the k-tuples of grammars are sentences of the system $\hat{G}^i$ if and only if a k-tuple of production rules $\langle p_1, p_2, \ldots, p_k)\rangle$ belonging to R is applied at every derivational step. $L(\hat{G}^i)$, which is the language generated by the system $\hat{G}^i$, is thus included in $L(G_1^i) \times L(G_2^i) \times \ldots \times L(G_k^i)$. We introduce a convention by which a symbol under an arrow denotes the set of the elements of V occuring in all the strings which can be obtained by a single application of a production rule of P to that symbol,* and we assume that $L(\hat{G}^i)$ is a set of k-tuples of sentences $\langle y_1, y_2, \ldots, y_k\rangle$ (where $y_j \in L(G_j^i)$), for which certain conditions are satisfied, namely:

$$L(\hat{G}^i) = \{\langle y_1, y_2, \ldots, y_k\rangle \mid y_j = \langle A_1^j, A_2^j, \ldots, A_{m_j}^j\rangle\,, \quad A_1^j \in \vec{C}^i\,, \quad A_{s+1}^j \in \overrightarrow{A_s^j}\}\,,$$

where $s = 2, \ldots, m_j$; $j = 1, \ldots, k$; $A_{m_j}^j \in T$; $A_r^j \in V - T$ (for $r = 1, \ldots, m_j - 1$).

We say that a string over T is a sentence of the relational grammar if and only if it has a phrase marker in this grammar. And according to the definition of a phrase marker of a sentence generated by a relational phrase structure grammar,** Q is a phrase marker of a sentence a belonging to the language $L(\langle\hat{Z}, C^i, \hat{G}^i\rangle)$ if and only if:

1° Q is a phrase marker of the sentence a in the grammar $\hat{Z}$,

2° for every sequence (branch) occurring in Q which is a sentence belonging to one (or some) of the languages $L(G_j^i)$, there must occur a k-tuple of sequences generated by the corresponding system $\hat{G}^i$ (or one of these), this sequence being an element of the k-tuple.

Let us now specify a relational phrase structure grammar for the subset of Mohawk sentences which contain contituents with cross dependency structure, as described by Postal.

A relational grammar for infinitely many Mohawk constructions in question is the system $\langle\hat{Z}, C^i, \hat{G}^i\rangle$, where $i = 1$; we shall now specify the three components of this system:

$\hat{Z}$ is simple phrase structure grammar containing among others the following production rules:

$$\mathrm{VP}_{com} \rightarrow \mathrm{Verb\ NP}$$

$$\mathrm{Verb} \rightarrow \mathrm{Pron_{pre}V_{base}V_{su}}$$

$$\mathrm{V_{base}} \rightarrow \begin{Bmatrix} \mathrm{V_{st}} \\ \mathrm{Nom_{ht}V_{base}} \end{Bmatrix}$$

$$\mathrm{Nom_{st}} \rightarrow \begin{Bmatrix} \mathrm{V_{st}Nom_{su}} \\ \mathrm{N_{st}} \end{Bmatrix}$$

* A precise definition is given in [8].

** A precise definition of a phrase marker of a sentence generated by a relational grammar is given in the *Note* at the end of this paper.

$$N_{st} \to \begin{Bmatrix} N_{st_1} \\ N_{st_2} \\ \vdots \\ N_{st_n} \end{Bmatrix}$$

$$V_{st} \to \begin{Bmatrix} V_{st_1} \\ V_{st_2} \\ \vdots \\ V_{st_m} \end{Bmatrix}$$

$$\mathrm{NP} \to \mathrm{Mod}\ \mathrm{N}$$

$$\mathrm{Mod} \to \mathrm{kik}$$

$$\mathrm{N} \to \mathrm{N}_{pre}\ \mathrm{N}_{base}\ \mathrm{N}_{su}$$

$$\mathrm{N}_{base} \to \begin{Bmatrix} \mathrm{Nom}_{st}\ \mathrm{N}_{base} \\ \mathrm{Nom}_{st} \end{Bmatrix}$$

where the symbols stand for the following grammatical categories: VP_{com} – complex verb phrase, Mod – modifier, V_{base} – verb base, V_{st} – verb stem, Pron_{pre} – pronominal prefix, V_{su} – verb sufix, N_{base} – noun base, Nom_{st} – complex nominalized stem, Nom_{su} – nominalizing suffix, N_{pre} – nominal prefix, N_{su} – noun suffix.

The second component of the relational grammar, that is, C is the symbol VP_{com}. The third component, $\hat{G}$, is as follows:

$$\hat{G} = \langle\langle G_1, G_2\rangle\ R\rangle\ ,$$

where G_1 is a simple phrase structure grammar with the rules:

(1) $\quad \mathrm{S}' \to \mathrm{Verb}\ \mathrm{V}_{base}\ \mathrm{Nom}_{st}\ \mathrm{N}_{st}\ \mathrm{N}_{st_i} \quad (\text{for } i = 1, \ldots, n)$,

(2) $\quad \mathrm{S}' \to \mathrm{Verb}\ \mathrm{V}_{base}\ \mathrm{Nom}_{st}\ \mathrm{V}_{st}\ \mathrm{V}_{st} \quad (\text{for } j = 1, \ldots, m)$,

(3) $\quad \mathrm{V}_{base} \to \mathrm{V}_{base}\ \mathrm{V}_{base}$;

G_2 is a simple phrase structure grammar with the rules:

(1) $\quad \mathrm{S}' \to \mathrm{NP}\ \mathrm{N}\ \mathrm{N}_{base}\ \mathrm{Nom}_{st}\ \mathrm{N}_{st}\ \mathrm{N}_{st_i} \quad (\text{for } i = 1, \ldots, n)$,

(2) $\quad \mathrm{S}' \to \mathrm{NP}\ \mathrm{N}\ \mathrm{N}_{base}\ \mathrm{Nom}_{st}\ \mathrm{V}_{st}\ \mathrm{V}_{st_j} \quad (\text{for } j = 1, \ldots, m)$,

(3) $\quad \mathrm{N}_{base} \to \mathrm{N}_{base}\ \mathrm{N}_{base}$.

In fact, the first two rules of G_1 and G_2 are presented in an abbreviated notation for n and m rules respectively (n being the number of noun stems, and m being the number of verb stems in Mohawk)*

* Notice that the number of noun stems and verb stems is finite, and it is only the number of complex nominalized stems which is infinite.

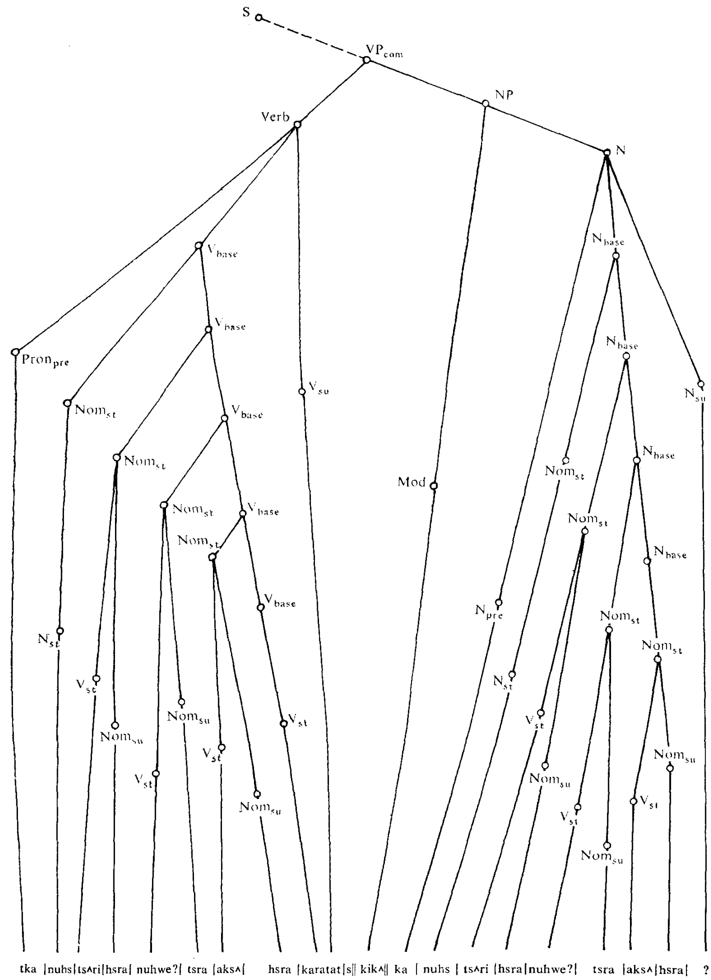

Fig. 1. Phrase marker of a sentence which belongs to the language $L(\langle \hat{Z}, C, \hat{G} \rangle)$; the English translation is: "She praises the evil of the liking of the finding of the house".

 R is a coupling relation containing the following pairs:

rule (1) of G_1 and rule (1) of G_2 (actually there will be n such pairs in accordance with the number of noun stems); rule (2) of G_1 and rule (2) of G_2 (there will be m such paris in accordance with the number of verb stems); finally there is a pair consisting of rule (3) of G_1 and rule (3) of G_2.

Accordingly, R is a one-one correspondence included in $P_1 \times P_2$, where P_1 and P_2 are the sets of production rules of G_1 and G_2, respectively, and it contains $(n + m + 1)$ pairs.

In accordance with the definition, a sentence generated by the grammar $\hat{Z}$ will belong to the language generated by the relational grammar, if and only if it has a phrase marker in the relational grammar. The conditions specified above guarantee

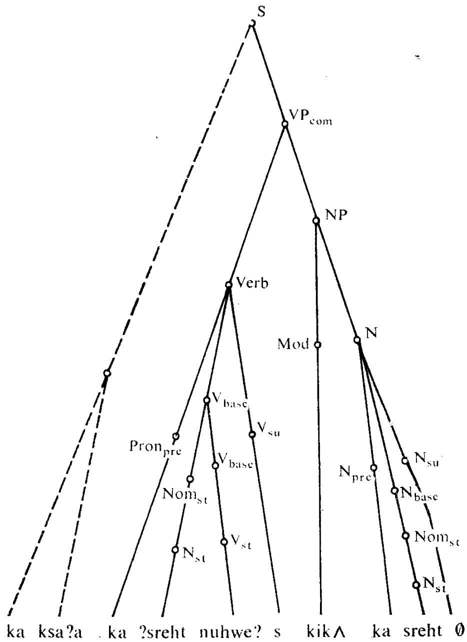

Fig. 2. Phrase marker of a sentence which belongs to the language $L(\langle \hat{Z}, C, \hat{G} \rangle)$; the English translation is: "The girl likes this car"; a rough, morpheme-by-morpheme translation is: "The girl car-likes this car".

that the cross dependencies** will be preserved between any of the infinite number of the complex nominalized stems in the verb base and in the noun base (see Fig. 1 and Fig. 2), and that non-sentences will be excluded. 271

For instance the two sequences:

kaksa?a kanuhwe?s kik∧ k?sreht

kaksa?a ka?srehthuhwe?s kik∧ kanuhsa?

which in a rough morpheme-by-morpheme translation would mean: *the girl house-likes this car*, and *the girl car-likes this house*, respectively, will not be accepted by

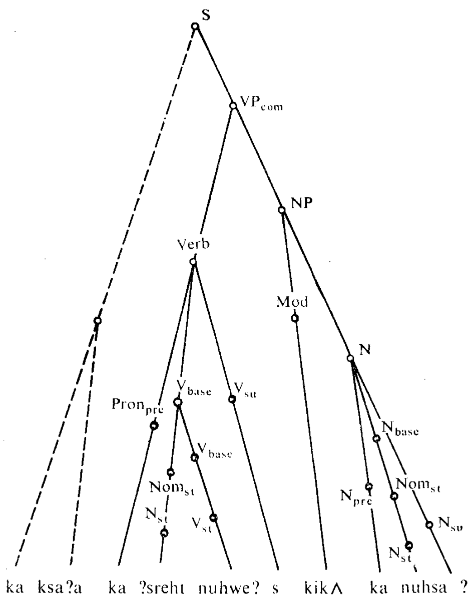

Fig. 3. Phrase marker of a non-sentence which belongs to the language $L(\hat{Z})$, but does not belong to the language $L(\langle \hat{Z}, C, \hat{G} \rangle)$; a rough, morpheme-by-morpheme translation is: "The girl car-likes this house".

** It is clear that the grammar of Mohawk constructions with cross dependencies is presented here as a system per se; accordingly, the occurrence of the modifier is obligatory in each derivation of this grammar; evidently, the rules would be different if they appeared as an integral part of the entire grammar of Mohawk.

our relational grammar, although they will both have a phrase marker in the grammar $\hat{Z}$. The phrase marker of the latter is given in Fig. 3; notice that in that phrase marker there occurs the sequence (branch of the tree): Verb $\mathrm{V_{base}}$ $\mathrm{Nom_{st}}$ $\mathrm{N_{st}}$ $\mathrm{N_{st_i}}$ (where $\mathrm{N_{st_i}}$ is the noun stem: *sreht*), but there occurs no sequence NP N $\mathrm{N_{base}}$ $\mathrm{Nom_{st}}$ $\mathrm{N_{st}}$ $\mathrm{N_{st_i}}$ (where $\mathrm{N_{st_i}}$ denotes the same noun stem, that is, *sreht*); the two sequences constitute a pair generated by the system $\hat{G}$, and if one occurs, the other one should occur too. Accordingly, the first condition of the definition (see p. 267) is satisfied, whereas the second one is not, and the sequence has no phrase marker in the relational grammar.

On the other hand, the phrase markers of the sentences presented in Fig. 1 and 2 belong to both the grammar $\hat{Z}$ and the relational grammar $\langle \hat{Z}, C, \hat{G} \rangle$, as clearly for all the sequences which are generated by G_1, we can find the corresponding sequences which are generated by G_2, and which constitute together the pairs of the system $\hat{G}$. Accordingly, both conditions of the definition are satisfied.

Thus we have shown that a relational grammar can enumerate the infinite set of Mohawk sentences, the structure of which was described by Postal. Moreover, it assigns to each sentence a phrase structure marker, whose linguistic meaningfulness can hardly raise any doubts on the part of the grammarian, whether traditional or not.*

Note: A definition of a phrase marker of a sentence generated by a relational grammar is given in [8] on page 520. However, on account of some errors, the editor of "Information and Control" has agreed to publish a letter to the editor with a corrected version of that definition in a forthcoming issue. The corrected version is as follows:

A phrase marker of the sentence a generated by a relational grammar $\langle \hat{Z}, C^i, \hat{G}^i \rangle$ is a phrase marker of the sentence a generated by the grammar $\hat{Z}$, such that satisfies the following conditions:

For every system of vertices $n, n_1^j, n_2^j, \ldots, n_{m_j}^j$ ($m \geqq 2$, $j = 1, \ldots, p$, where p is the number of all the sequences of vertices ordered by the relation G_1 defined in 1.2.d),

if $n_1^j \in \vec{n}$, $n_{s+1}^j \in \overrightarrow{n_s^j}$ $(s = 2, \ldots, m_j)$, $G_2(n_1^{j-1}, n_1^j)$, where G_2 is a weak relation of linear order defined in 1.2.d, if $\varphi(n) = C^i$ and if any sequence $\varphi(n_1^j)\,\varphi(n_2^j) \ldots \varphi(n_{m_j}^j)$ belongs to $L(G_r^i)$ $(1 \leqq r \leqq k)$,

then a k-tuple of sequences $\langle \varphi(n_1^{j_1})\,\varphi(n_2^{j_1}) \ldots \varphi(n_{m_{j_1}}^{j_1}), \varphi(n_1^{j_2})\,\varphi(n_2^{j_2}) \ldots \varphi(n_{m_{j_2}}^{j_2}), \ldots, \ldots, \varphi(n_1^{j_k})\,\varphi(n_2^{j_k}) \ldots \varphi(n_{m_{j_k}}^{j_k}) \rangle$ belongs to $L(\hat{G}^i)$ $(j_r = j$, and $1 \leqq j_i \leqq p)$.

(Received September 10th, 1965.)

* I am indebted to Professor P. M. Postal for his comments on the structure of Mohawk sentences.

REFERENCES

[1] N. Chomsky: Syntactic Structures. Mouton and Co., The Hague 1957.
[2] N. Chomsky: Three Models for the Description of Language. IRE Transactions on Information Theory, Vol. *IT-2* (1956).
[3] N. Chomsky: On Certain Formal Properties of Grammars. Information and Control *2* (1959), No. 2.
[4] N. Chomsky: Formal Properties of Grammars. Handbook of Mathematical Psychology II. Ed. E. Galanter, R. D. Luce, R. R. Busch. John Wiley, New York 1964.
[5] N. Chomsky, G. Miller: Introduction to Formal Analysis of Natural Languages. Handbook of Mathematical Psychology II. Ed. E. Galanter, R. D. Luce, R. R. Busch, New York 1964.
[6] P. M. Postal: Constituent Structure. Mouton and Co., The Hague 1964.
[7] P. M. Postal: Limitations of Phrase Structure Grammars. The Structure of Language. Ed. J. J. Fodor, J. J. Katz, 1964.
[8] I. Bellert: A. Relational Phrase Structure Grammar and Its Tentative Applications. Information and Control *8* (1965), No. 5.

VÝTAH

Relační frázová gramatika a její použití pro konstrukce jazyka Mohawků

IRENA BELLERT

Autorka uvádí definici relační frázové gramatiky, kterou již popsala ve své dřívější práci [8], a její použití na určité konstrukce jazyka Mohawků, jednoho ze zachovaných severoirokézských jazyků. Konstrukce jazyka Mohawků již diskutoval P. M. Postal [7], který ukázal, že se jazyk Mohawků vymyká bezkontextovému popisu.

V práci se ukazuje, že relační frázová gramatika je schopná vyčíslit nekonečnou množinu konstrukcí jazyka Mohawků se vzájemnými závislostmi a přiřadit pro každou z nich frázový ukazatel.

Dr. Irena Bellert, Department of General Linguistics, Warsaw University, Warsaw. Poland.