[*International Conference on Machine Translation of Languages and Applied Language Analysis*,
National Physical Laboratory, Teddington, UK, 5-8 September 1961]

A NOTE ON CATEGORIAL GRAMMARS

by

R.P. MITCHELL

(Lockheed Missiles and Space Company,
California, U.S.A.)

THIS paper presents a technique for studying the structure and theory of categorial grammars. Grammars of the type studied by Y. Bar-Hillel and others are shown to be representable over a two-symbol alphabet. A trivial corollary is that the category "sentence" is decidable in all these grammars. A decision problem for normal categorial grammars, of which restricted categorial grammars are an example, is shown to be recursively undecldable.

1. PRELIMINARY DEFINITIONS

Denote the set of natural numbers by I, and the set of all n-tuples over I by S. Define $S_{k,n} = \{(x_1, \dots, x_n) | x_i < k, 1 \leq i \leq n\}$ for all n. It is convenient to abbreviate $(x_1, \dots, x_a)$ to $x^{(a)}$. As usual, two n-tuples are said to be equal iff their corresponding elements are equal:

$$x^{(n)} = y^{(n)} \iff \bigwedge_{i=1}^{n} x_i = y_i$$

We shall also find it convenient to use the notation $m.x^{(a)}$, m a natural number, to mean the a-tuple $(m.x_1, \dots, m.x_a)$.

We define the set A of elements of S by the condition, for all n,

$$A = \{(x^{(n)}, y^{(n)}) | x^{(n)} \neq y^{(n)}\},$$

and we use Davis' [4] definition of the characteristic function. Thus,

$$C_A(x^{(n)}, y^{(n)}) = \begin{cases} 0, & \text{if } x^{(n)} \neq y^{(n)} \\ 1, & \text{if } x^{(n)} = y^{(n)} \end{cases}$$

Denote $\bigcup_n S_{m,n}$ by S_m. We define a function $F_m : S_m \to I$ having the property that, for each $x = x^{(n)} \epsilon S_m$ for some n, $F_m(x) = x_1 + m.x_2 + \ldots + m^{n-1}.x_n$. Observe that $F_1(x) = 0$ and F_0 is not defined.

F_m is not 1-1 as defined; however, if we define an equivalence relation $\sim$ such that, if x and y are elements of S_m, and if $x = (x_1, \ldots, x_n)$ has the property that $x_n \neq 0$, and $y = (x_1, \ldots, x_n, 0, \ldots, 0)$, then $x \sim y$, and define F_m over such equivalence classes, then F_m is 1-1. For our purposes this is not essential. F_m is recursive since it can be defined by composition in terms of m^x, which is recursive, and sums and products. Define the recursion equation by the function $H : I^2 \times S_m \to I$, for $x \epsilon S_m$, as follows

$$H(0, m, x) = x_1,$$
$$H(z+1, m, x) = x_{z+2} \cdot m^{z+1} + H(z, m, x)$$

Then $F_m(x) = H(n-2, m, x)$.

We define a function $K_m : S_m \times S_m \to S_m$ having the following properties for any pair of elements $x = x^{(a)}$, $y = y^{(b)}$ of S_m:

$$K_m(x,y) = \begin{cases} x^{(a-b)} \cdot C_A[(x_{a-b+1}, \ldots, x_a), y^{(b)}], & \text{if } a > b; \\ (y_{a+1}, \ldots, y_b) \cdot C_A(x^{(a)}, y^{(a)}), & \text{if } a < b; \\ x^{(a)} \cdot C_A(x^{(a)}, y^{(a)}), & \text{if } a = b. \end{cases}$$

Finally, we also require the function $L_m : S_m \times S_m \to S_m$ such that, if $x^{(a)} \epsilon S_m$, $y^{(b)} \epsilon S_m$,

$L_m(x^{(a)}, y^{(b)}) = (x_1, \dots, x_{a-1}, y_2, \dots, y_b) \cdot C_A(x_a, y_1)$

K_m and L_m are both decidable functions. To show this, define the set $B = \{(x,y) \mid x < m,\ y < m,\ x \neq y\}$ for all x, y and some m. Then $C_A(x^{(n)}, y^{(n)}) = C_B(x_1,y_1) \cdot C_B(x_2,y_2) \cdot \dots \cdot C_B(x_n,y_n)$, which is recursive.

2. CHARACTERIZATIONS OF CATEGORIAL GRAMMARS

In [1], Bar-Hillel, *et al.*, define three types of categorial grammars. Our main result in this section is the fact that these grammars are examples of a large class of grammars obtainable from a general theory. We shall give an example of another type of categorial grammar also obtainable.

As in the previous section, $S_{m,n}$ denotes the set $\{(x_1, \dots, x_n) \mid x_1 < m,\ 1 \leq 1 \leq n\}$. We define *the string corresponding to the n-tuple* $x^{(n)} \epsilon S_{m,n}$ as the concatenate of the symbols $x_1, x_2, \dots, x_n$ in the order given by the n-tuple, and denote this string by $x_{(n)}$; thus, the string corresponding to the pair $(x^{(n)}, y^{(m)})$ is the concatenate of the strings $x_{(n)}, y_{(m)}$, viz., the string $x_1 \dots x_n\, y_1 \dots y_m$. We shall find it convenient to write the arguments of K_m as strings rather than n-tuples. Let $\sigma_{m,n}$ be the set of strings corresponding to the elements of $S_{m,n}$, and σ_m denote the set $\bigcup_n \sigma_{m,n}$.

A bidirectional categorial system (BCS) is defined in [1] as an infinite set of symbols $\mathbf{C}$ obtained from a given finite set $\mathbf{C}_p$ in the following manner;

(1) If $x_1 \epsilon \mathbf{C}_p$, then $x_1 \epsilon \mathbf{C}$;

(2) If x, $y \epsilon \mathbf{C}$, then $[x/y] \epsilon \mathbf{C}$;

(3) If x, $y \epsilon \mathbf{C}$, then $[x \backslash y] \epsilon \mathbf{C}$.

Following Bar-Hillel, we shall call the elements of $\mathbf{C}_p$ *primitive categories* and the elements of $\mathbf{C}$ *categories*.

Given any BCS, there is a 1-1 correspondence between the elements of C_p and the elements of the set $\{1, \ldots, m-2\} \subset \sigma_{m,1}$; if, further we use 0 for, say, $/$, and m-1 for $\backslash$, then σ_m is the set of all possible strings over the set $\sigma_{m,1}$, including this set. The set corresponding to C is in fact, a subset of $\bigcup_{k=1}^{\infty} \sigma_{m,2k-1}$. We next define a binary relation on σ_m such that the following conditions hold for all x and $y \in \sigma_m$:

$Ex \ldots x, x$; $Ex0, x$; $E(m-1)x, x$; $E0x, 0$; $Ex(m-1), 0$; $Ex0y, x0y$; $Ex(m-1)y, x(m-1)y$.

The basic reason for such a relation is to ensure that application of a cancellation rule leads to a string which is a category in the grammar. The first condition, $Ex \ldots x, x$, is not essential for this purpose, but strings of the form $x \ldots x$ do not occur in the grammars being considered here, and it is convenient to consider such strings equivalent to the single category x.

The grammars defined in [1] have no rule for cancellation of sequences of the form x, x; nor for cancellation of sequences of the forms x/y, y/z and x\y, y\z. If we wish to characterise cancellation in a BCG by K_m, we must define $K_m(x^{(a)}, y^{(a)}) = (\overbrace{0, \ldots, 0}^{a})$, which is to say, strings containing the same number of primitive categories never cancel. Grammars of the type considered by Lambek [3], similar to categorial grammars as defined by Bar-Hillel [1] in certain other respects, do have a rule of the form

$$x/y,\ y/z \rightarrow x/z\ ; \qquad x\backslash y,\ y\backslash z \rightarrow x\backslash z.$$

We can see no particular advantage of such a rule for these grammars, and in fact elimination of this rule simplifies our intended characterization by asserting also that x, x does not cancel. Accordingly, we define a new function K'_m such that $K'_m(x^{(a)}, y^{(b)}) = K_m(x^{(a)}, y^{(b)})$ whenever $a \neq b$, and $K'_m(x^{(a)}, y^{(b)}) = (\overbrace{0, \ldots, 0}^{a})$ when a = b. Then we say that a sequence α of strings *directly cancels** to a sequence β iff

$$\alpha = \gamma,\ x^{(a)},\ y^{(b)},\ \delta \quad \text{and} \quad \beta = \gamma,\ EK'_m(x^{(a)}, y^{(b)}),\ z^{(c)},\ \delta$$

for some γ and δ, and $z^{(c)} \neq 0$.

---

* The terminology is Bar-Hillel's [1].

To illustrate the notions defined, take the BCG consisting of a finite vocabulary V, an assignment function **A**, a BCS **C** whose $\mathbf{C}_\rho = \{n, s\}$, and in which s is the distinguished element. We take the set $\sigma_{4,1}$ and map $/ \to 0$, $n \to 1$, $s \to 2$, and $\backslash \to 3$. The sequences

(a) 101, 1, 13201, 1 ;

(b) 101, 101, 1, 132, 232

cancel in the following steps:

(a) $K'_4(101, 1) = 10$; E10, 1; $K'_4(1, 13201) = 3201$;

E3201,201; $K'_4(201, 1) = 20$; E20,2; 2.

(b) $K'_4(101, 101) = 000$; E000,0; $K'_4(101, 1) = 10$;

E10,1; $K'_4(101, 1) = 10$; E10,1; $K'_4(1, 132) = 32$;

E32,2; $K'_4(2, 232) = 32$; E32,2; 2.

The following string does not cancel:

(c) 1, 101, 13201, 1

since $K'_4(1, 101) = 01$ but E01,0.

Consider next the question, "What is the simplest grammar we can define using a $\sigma_{m,n}$ and some cancellation pair (K_m, E) defined on σ_m?" Note that $\sigma_{2,2}$ is the smallest set in a domain for which cancellation is defined. We therefore take the set $\{10, 01, 11\}$ as a primitive category set, and no distinguished element. We use simple "equivalence": Ex, x, for all x, and the function L_2 as a cancellation pair. A sequence α is said to cancel to a sequence β iff

$$\alpha = x^{(2)},\ y^{(2)},\ \delta \quad \text{and} \quad \beta = z^{(2)}\ ,\ \delta$$

for some δ, and $F_2(z^{(2)}) \neq 0$; i.e., left to right cancellation. Finally, we assume a finite vocabulary V and an assignment function $\mathbf{A} : V \to \{10, 01, 11\}$.

We observe that the sequence 10, 01 cancels, while the sequence 01, 10 does not cancel. We may therefore take 10 as a "nominal" category and 01 as a "verb" category. The remaining string, 11, we use as a "catch-all" category. Next, we note that the following sequences all cancel:

$$11,\ 10 : L_2(11,\ 10) = 10;$$

$$11,\ 11 : L_2(11,\ 11) = 11;$$

$$10,\ 01,\ 10 : L_2(10,\ 01) = 11, \quad L_2(11,\ 10) = 10;$$

$$10,\ 01,\ 11 : L_2(10,\ 01) = 11, \ L_2(11,\ 11) = 11.$$

By the first sequence, 11 includes pre-nominals; from the second, all pairs of words in 11 behave as a word in 11; by the third sequence, transitive verbs as well as intransitive are in 01; and by the last sequence, 11 includes post-verbal modifiers as well as pre-nominals. On the other hand, the sequence 11, 01 shows that 11 does not include pre-verbal modifiers. Hence, the grammar, coarse though its categories are, does have limitations,

3. A DECISION PROBLEM

The examples given in the preceding section illustrate the main advantage of our characterization: Its flexible and consistent notation permit a range of experimentation in grammars of fixed vocabulary and different image sets for the assignment function. Proceeding in a manner similar to that used in constructing the "minimal" grammar on $\sigma_{2,2}$, we can construct a class of grammars over σ_2, keeping the vocabulary fixed, but changing the assignment function and the cancellation pair (K, E) as required.

Consider now the set $\sigma_{m,1}$, $m \geq 2$. Then for some $x \in S_2$ and $y \in \sigma_{m,1}$, the equation $y = F_2(x)$ holds; in fact, there is a denumerably infinite set of such elements. We take the first element $\overset{*}{x}$ such that $F_2(\overset{*}{x}) = y$, for each $y \in \sigma_{m,1}$. Writing $x^*_{(n)}$ for the string corresponding to x^*, we obtain a set of strings σ^*_m from σ_m by substituting for each occurrence of y in a string of σ_m the corresponding x^*, deleting recurrences of the same strings. Then σ^*_m is a set of strings of 0's and 1's, and is the same set as σ_2 : $\sigma^*_m = \sigma_2$. If we have an assignment function $\mathbf{A} : V \rightarrow \sigma_{m,k}$, we can map V into a set $\sigma_{2,n}$ as well; since, further, K, K', and L were defined for all m, we need only K_2, K'_2, and L_2. Thus every categorial grammar of the types considered here is representable over strings of a two-symbol alphabet {0, 1}.

For example the categories of the BCG with $C_p = \{n, s\}$, considered previously, may be obtained as strings formed from the set $\sigma_{2,2}$, by writing 00 for / , 01 for n, 10 for s, and 11 for \; with corresponding modifications in E, the same cancellation rules hold using K'_2 instead of K'_4. It is, of course, quite immaterial which of the sets $\sigma_{2,2}, \sigma_{2,3} \ldots, \sigma_{2,n}$ we use to obtain primitive categories, provided only, given a procedure for obtaining all categories of a grammar, we can effectively determine when a given pair of strings cancel to a string belonging to the set of categories. Since K_2, K'_2 and L_2 are recursive, the only remaining question is whether we can effectively determine when an arbitrary string is a category in a given grammar. This is decidable if E is a recursive relation. For the grammars we have considered, E is clearly recursive, being of the forms $E\alpha\beta, \alpha$, $E\alpha\beta, \beta$, $E\alpha \ldots \alpha, \alpha$, for α, β arbitrary (possibly null) strings. Thus, for example, cancellation of a sequence of strings to a distinguished category is decidable.

We shall call a categorial grammar *normal* iff the following are satisfied:

(i) V is finite

(ii) $\mathbf{A} : V \to \sigma_2$, i.e., the assignment function takes the vocabulary onto a subset of σ_2.

(iii) The set of categories is the set of assertions of a normal system* on {0,1}.

(iv) A distinguished category α_0.

---

* For the definition of a normal system as used here, see Post [5,6].

An example of a normal categorial grammar is the restricted categorial grammar of Bar-Hillel [1]. Using the notation of Bar-Hillel [1], with the understanding that $\Delta_1, \ldots, \Delta_m$ are distinct strings over {0, 1}, and Δ_1 is the initial string, we have the rules:

$$\alpha \Delta_1 \rightarrow \Delta_i \backslash \Delta_j$$

$$\alpha \Delta_i \backslash \Delta_j \rightarrow \Delta_1 \backslash \Delta_j \backslash \Delta_k$$

where α is any string (possibly null). Hence, every category in a RCG is an assertion in a normal categorial grammar.

A decision problem known to be recursively unsolvable is that of determining, for an arbitrary normal system, whether an arbitrary (non-null) string belonging to σ_2 is an assertion of the normal system [5, 6]. We thus have the result:

The decision problem of determining, for an arbitrary normal categorial grammar belonging to the class of such grammars over a finite vocabulary, whether an arbitrary string belonging to σ_2 is a category in that grammar is recursively unsolvable.

REFERENCES

1. BAR-HILLEL, Y., GAIFMAN, C., and SHAMIR, E., "On Categorial and Phrase-structure Grammars", *Bull. of Research Council of Israel,* 1960 **9F**,No.1.
2. BAR-HILLEL, Y., PERLES, M., and SHAMIR, E., "On Formal Properties of Simple Phrase-structure Grammars", *Tech. Rpt. No. 4,* Applied Logic Branch, Hebrew University of Jerusalem, 1960.
3. LAMBEK, J., "The Mathematics of Sentence Structure", *Amer.Math. Monthly,* 1958, **65:3** , 154-170.
4. DAVIS, M., Computability and Unsolvability, New York: 1958, McGraw-Hill Book Co.
5. POST, E., "Recursively Enumerable Sets of Positive Integers and Their Decision Problems", *Bull. Amer. Math. Soc.,* 1944. **50**, pp. 284-316.
6. POST, E., "Variant of a Recursively Unsolvable Problem", *Bull. Amer. Math. Soc.,* 1946, **52**, pp. 264-268.