

XVII. MECHANICAL TRANSLATION*

V. H. Yngve
J. R. Applegate
J. L. Bennett
J. S. Bross
Elinor K. Charney

A. N. Chomsky
D. A. Dinneen
Muriel Kannel
E. S. Klima
K. C. Knowlton
D. Lieberman

G. H. Matthews
A. B. Phillips
Rosemarie Sträussnigg
J. J. Viertel
Ruth V. Ward

A. TRANSLATION STUDIES. I.

Communication between speakers of different dialects of the same language shares features with communication between speakers of different languages. In both cases, some translation is necessary. In communication between speakers of the same language, structural similarities may account for the fact that translation is less complex and may be carried on by the hearer. As structural differences increase, the process becomes more complex until an intermediary must be used. In this study an attempt was made to determine the points of structural divergence that have the greatest effect on communication.

The subjects of this case study were three brothers who were four, five and one-half, and eight and one-half years old when the study began. They are the sons of native-born parents whose speech is typical of the small New England community near Boston where they live. Deviations from the speech of the adult community were readily apparent in the speech of all three children. The two younger children spoke a dialect that was almost completely unintelligible to adults (even, at times, to their parents). Yet they were able to communicate freely with each other and with their playmates, and with their older brother. He spoke a dialect that could be clearly recognized as a form of the community dialect with various distortions. The most obvious of these was severe stuttering; another was the reversal of the order of syntactic constituents.

A description of the younger children's dialect shows that the difference between it and the adults' dialect can be traced to two phonological rules found in the children's dialect but not in that of the adults. Application of the first of these rules converts all fricatives to stops; and application of the second converts any stop repeated within word boundaries to a glottal stop. These two rules, when applied in the proper order (application of the second must precede application of the first), are sufficient to explain the differences between the two dialects. The apparent differences in morphology can be traced to these rules.

A possible reason for the preservation of a deviant form of speech by the younger children was that the oldest child, who understood them but whose speech was closer to that of the adult community, was willing to act as a translator for them. Usually, a

* This work is supported in part by National Science Foundation.

(XVII. MECHANICAL TRANSLATION)

child's speech is considered a deviant form of adults' language, and he is required to alter it until it conforms to the pattern of the adults' language. This is necessary if he is to be understood. This was not true in this case, however. The fact that the oldest child was able to serve as a translator when communication between the younger children and adults broke down minimized the pressure to alter their speech so that they would be understood.

The translation of the younger children's utterances required only the elimination of the two special phonological rules that were peculiar to their speech. Apparently, the oldest child understood this. When he translated, however, the deviations that were characteristic of his speech in free conversation disappeared. There was no stuttering and no reversal of sentence constituents. This was true only when the translation was given as a direct quotation ("He said, 'I want some candy!' "). When the translation was changed to an indirect quotation ("He said that he wanted some candy"), the deviations returned.

These facts suggest that the three-stage translation model, which consists of (a) analysis of input sentence to obtain a structural specifier, (b) conversion of this specifier to a corresponding one for the output sentence, and (c) construction of the output sentence, may not be entirely accurate. In this case study, the generation of the output sentence by the oldest child proceeded in quite a different way from the usual generation of sentences. If it had not, the translations would have shown the deviations that were characteristic of his speech in other situations. Other studies of translation may lead to a better description of the translation process itself.

J. R. Applegate

B. RECOGNITION OF THE STRUCTURE OF SENTENCES

The grammar of a natural language can be thought of as a finite list of ordered rules which is such that the application of any particular set of these rules produces a sentence of the language. Some of the rules are optional recursive rules; that is, there is no limitation on the number of times that they may be applied in the production of a sentence. Thus, there is no limit on the number of different sentences that a grammar can produce. Other rules are optional but not recursive; that is, they may be applied only once. Still other rules are obligatory; they must be applied in the production of every sentence. A grammar of this type can be viewed as a certain function that has a variable for each optional rule of the language, and the order of these variables is the same as the order of the rules. The values of the variables are numerals representing the number of times that the corresponding rule is applied. There are some sets of rules in the grammar, all the members of which are optional, but no more than one of which can be used in the production of any one sentence. In such cases, a single variable can

represent the whole set of rules, and its value indicates which rule of the set is to be applied. In this way a one-to-one correspondence between the combinations of values—which will hereafter be called "specifiers"—and the sentences of the language is established, and thus the sentences are denumerable.

In addition, the sentences of a language can be placed in a well-ordering without accumulation points. This ordering is given to the specifiers by assigning a number to each specifier in the following way. All of the possible combinations of recursive rules of the grammar are well-ordered by the diagonal method, and these combinations are mapped onto the positive integers. The positive integer that corresponds to the combination of recursive rules in the specifier is placed in the leftmost place in the number assigned to that specifier. Those values in the specifier that refer to nonrecursive rules are placed to the right of this, in the same order that they have in the specifier. There is then a one-to-one correspondence between specifiers and their assigned numbers. Each place in these numbers has a different base, but only the leftmost place has infinity for its base. These numbers can be put in numerical order—and therefore the specifiers assigned to them—and hence the sentences corresponding to the specifiers can be given the same order.

Let us return to the examination of the grammar rules. Nearly all of the rules in a grammar have the characteristic of either keeping constant or increasing the number of symbols in the sentence that is being produced. A few of the rules decrease the number of symbols by deleting some of them. For each of these rules that delete symbols there is a maximum number of symbols that can be deleted. Furthermore, there is only a small number of combinations of these rules that can be applied in the production of a sentence. This means that, given a sentence that is n symbols long, a constant k can be calculated with the characteristic that, in the production of the sentence, before the application of those rules that delete symbols, there will be no more than $(n+k)$ symbols in the sentence.

Since grammar rules (with the exception of the deletion rules) either maintain the number of symbols in the sentence or increase their number, and since, in the numerical values that are given to grammar rules, a lower numeral represents a fewer number of times that the rule is applied, we conclude that, in general, the order of sentences is from shorter to longer. Furthermore, this order is roughly equivalent to the intuitive notion of from less complex to more complex because, in general, the complexity of a sentence increases with the number of times a recursive rule is applied in its production, for it is by such rules that clauses are added to a sentence and the relationship between clauses established. Therefore, given any number m , it is possible to calculate a specifier in the order of specifiers which is such that there will be no specifier following it that corresponds to a sentence of m symbols or less. Thus, the grammar of a natural language is a decidable theory. To decide whether a given string of symbols is a sentence

(XVII. MECHANICAL TRANSLATION)

of a certain language, one calculates the maximum number of symbols there would have been if no deletion rules had been applied in its production and then produces all the sentences of the languages which have that maximum number of symbols, or less, using the order of the specifiers to be sure that every sentence of that set is actually produced just once. All of these sentences will be produced in a finite time because there are no accumulation points in the order of specifiers. If one of the sentences matches symbol-for-symbol the string in question, then that string is a sentence of the language. Another feature of this decision procedure is that for each possible derivation of a given string of symbols, the decision procedure will produce a sentence that matches the string, symbol-for-symbol, and, because of the particular relationship between the specifiers and the grammar of the language, the specifier used in producing each sentence will be an unambiguous representation of the derivation of the sentence—that is, the grammatical analysis of the sentence.

I shall now define the notion "linguistically significant class of sentences." A class of sentences is linguistically significant if, in producing the members of this class, the same grammar rule has been applied the same number of times to all of them. This is a generalization of the intuitive notion, "sentence type." The specifiers for the members of a linguistically significant class are alike in that the same numeral occurs in the place that corresponds to the particular defining rule; this numeral represents the number of times that the rule is applied. In addition, the logical product of two linguistically significant classes of sentences, if it is nonempty, is also a linguistically significant class, and the specifiers for such a class will have more than one place containing the same numeral. Note that if the product of two classes is empty, it would be because each class would have kept the same rule constant but at a different number of times. The corresponding class of specifiers would be a class all of whose members simultaneously have two values in the same place—an impossibility.

These considerations are of importance in the general problem of the recognition of sentences, whether it be for a mechanical recognition procedure, such as those procedures that are being investigated by various mechanical-translation and information-retrieval projects, or for an explanation of how human beings use language. The close relationship among the notion "linguistically significant class of sentence," the grammar of the language, and the decision procedure that automatically provides a proof for the sentence in question, suggests a method by which the recognition procedure might be carried out in a practical manner. The input sentence and its environment are examined for clues that would place it in some linguistically significant class of sentences. This class is easily defined by a specifier in which some of the variables are left unspecified. The decision procedure uses this partially specified specifier to discover all of those members of this class of sentences that match the input sentence symbol-for-symbol, and provides all of the grammatical analyses of the input sentence. Those sentences

(XVII. MECHANICAL TRANSLATION)

that are produced by the decision procedure but do not match the input sentence can be used to reduce the number of sentences of the class that still must, in fact, be produced, for each of them can define a class of sentences having the characteristic of overlapping with the class that is being produced but not of including the input sentence.

G. H. Matthews

