

XX. MECHANICAL TRANSLATION*

V. H. Yngve	W. S. Cooper	G. H. Matthews
J. R. Applegate	D. A. Dinneen	A. B. Phillips
J. S. Bross	E. S. Klima	Rosemarie Straeussnigg
Elinor K. Charney	K. C. Knowlton	J. Viertel
A. N. Chomsky	D. Lieberman	Ruth V. Ward

RESEARCH OBJECTIVES

The primary objective of this basic research program is to find out how languages can be translated by machine. Secondary objectives are concerned with evaluating the fidelity which can be achieved with different approaches, the usefulness of the translations for various purposes, and their respective costs. Another objective is to add to the general knowledge of noncomputational uses of digital computing machinery and to a basic understanding of human communication.

The Mechanical Translation group has always stressed a basic, long-range approach to translation. We are placing emphasis on completeness where completeness is possible, and on an attempt to find out how to do a complete job where it is not now possible. We are not looking for short-cut methods that might yield partially adequate translation at an early date — an important goal pursued by other groups. Instead we are looking for methods that will be capable of yielding fully adequate results whenever they apply. We are thus seeking definitive solutions that will constitute permanent advances in the field rather than ad hoc or temporary solutions that may eventually have to be discarded because they are not compatible with improved systems.

In order to write adequate translating routines, we need, among other things, an adequate and detailed knowledge of the languages in question — a knowledge of their formal properties as codes and a knowledge of how they are used to communicate. Linguistic research on the structure of individual languages thus constitutes an important part of our effort. German and English are being given primary attention, but other languages are also being studied. Each language is studied as an isolated system. The translation relationship between languages is a separate question and is being given separate consideration.

An important part of our objectives is to investigate the various questions raised in the attempt to mechanize language operations. We feel that some of the problems of mechanization can profitably be studied without reference to any particular language. There may be aspects of the manipulation of language that are essentially nonlinguistic.

One of our aims continues to be the development and provision of adequate research tools in the form of computer programs that will facilitate linguistic research by providing the linguist with easy access to the machine. Thus far, we have developed a method of searching large quantities of English or German text for examples of particular constructions. In cooperation with the Computation Center, M. I. T., we have developed and have nearly finished programming an automatic programming system that provides the linguist with a problem-oriented notation system for use in writing computer programs.

Some work is also being carried out that has as its aim the investigation of the feasibility of programming computers in English. This work is being carried out in cooperation with the Solid-State and Molecular Theory Group of Massachusetts Institute of Technology.

V. H. Yngve

A. THE GERMAN NOUN PHRASE

The work on the description of the German noun phrase continues. It has been extended to include the description of the semantic correlates of syntactic structures.

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

(XX. MECHANICAL TRANSLATION)

This problem can be characterized as follows.

The description of a specific structure can be presented in the form of an ordered set of rules, the application of which will permit the generation of the structure in question. At various points in the set of rules, a decision is made to apply one rule rather than another. This choice determines the selection of some of the later rules to be applied, or it may determine that at a later point in the set of rules, more choices will have to be made. The problem that is relevant for mechanical translation is: How can the factors determining these choices be characterized? As an example, consider the following case: With a noun stem as the head of the noun phrase, the adjectival modifiers may occur in a post-nominal relative clause or as individual words in prenominal position between the article and the noun. These two structures are the results of the application of different sets of rules in the construction of the noun phrase, that is, the sets of rules are identical up to a certain point, after which they diverge. The question to be answered is: What factors influence the choice of rules which in one case will place the modifier in a clause, but in the other case place it before the noun?

It should be possible to represent the factors affecting the choice of specific sets of rules that are to be applied in such a way that given the representation, a corresponding set of rules can be automatically selected from the total set of rules for producing grammatical structures; that is, the representation serves as a specifier for a given structure. The converse is also possible; that is, given a specific structure, it should be possible to determine, first, the set of rules applied to construct it, and second, the specifier that represents this set of rules. Since the process of translation is conceived as a three-step process: (a) analysis of a sentence in the source language to obtain a sentence specifier, (b) conversion of the source language specifier to the target language specifier, and (c) construction of the sentence in the target language from the specifier, a solution to the problems mentioned is essential to the preparation of a translation program.

Additional work on the manipulation of grammatical rules in human communication is also being done. Investigation of the processes involved in converting utterances produced by the application of one set of rules to those produced by the application of another set that differs by one or two rules, or in the order in which the rules are applied, will be useful in preparing programs for the conversion of sentence specifiers.

J. R. Applegate

B. THE ERGATIVE RELATION IN HIDATSA

The Siouan languages have often been cited as languages that exhibit the ergative relation. The ergative relation is said to exist in a language when the mark of the object of transitive verbs is the same as the mark of the subject of intransitive verbs. This

mark may be a case affix, an agreement morpheme, or merely a position in the sentence. In Siouan studies, this relation has usually been referred to as a distinction between actors and goals, in which goals function as objects of transitive verbs and subjects of most intransitive verbs, whereas actors function as subjects of transitive verbs and of a relatively few intransitive verbs. Early investigators have often referred also to the substantially passive nature of the Siouan languages. I have been able to show that, in viewing the sentences of Hidatsa (a Siouan language) in the context of the whole grammar, rather than as isolated events, the ergative relation is not a fundamental feature of Hidatsa, if it can be said to exist at all.

Hidatsa has two kinds of verb phrases which differ from each other in several respects, one of which is that they require their subjects to take different case morphemes. Thus

macé A cáha íkā (= the man saw a stick)
 mía A nīti (= the woman is hungry)
 macé G nūhí (= the man stood up)
 cáha G hīsi (= the stick is red)

where the first word of each sentence is the subject; A and G represent the case morphemes that are added to the subject, and are selected by the verb phrase of the sentence; and the rest of the sentence is the verb phrase.

There is a transformation in Hidatsa that substitutes a constituent sentence for the object noun of a matrix sentence on the condition that the verb of the matrix sentence belong to a certain class.

(i) $X - N - V_d - Y \rightarrow X - S - V_d - Y$

where V_d represents that class of verbs that may take sentences as their objects. Thus, the two sentences: macé G nūhí (= the man stood up) and mía A macé íkā (= the woman saw a man), in which macé (= man) in the second sentence is the object noun of the V_d -stem íkā (= see), can be transformed into mía A macé G nūhí íkā (= the woman saw the man stand up). One of the several V_d -stems is the stem hē (= use, cause). Thus, the sentences: macé A nī íkā (= the man saw you) and náka A ási hē (= the child used a spoon) can be transformed into náka A macé A nī íkā hē (= the child caused the man to see you).

The V_d -stem hē is unique, in that transforms of (i) that contain it as the main verb stem, undergo another transformation. This transformation removes the stem hē if it follows one of a large class of verb stems, which normally require their subjects to take the G case.

(ii) $X - V_{tr} - hē - Y \rightarrow X - V_{tr} - Y$

where V_{tr} represents that class of stems which requires the removal of a following hē. Thus, there are the sentences:

(XX. MECHANICAL TRANSLATION)

macé A ní A níti hē (= the man made you hungry)

macé A ní G nūhí hē (= the man made you stand up)

macé A cáha G hátha (= the man made the stick be broken/the man broke the stick)

where the third sentence comes from (i) and (ii) applied to the sentences macé A cáha hē (= the man used a stick) and cáha G hátha (= the stick is broken).

Consider the following sentences, of which the first and third are transforms of (i) and (ii) and the verb-stems hátha (= break) and kání (= ask for) are V_{tr} -stems:

náka A cáha G hátha (= the child broke the stick)

cáha G hísi (= the stick is red)

náka A macé G hūpá kání (= the child asked the man for a shoe)

macé G hinámi (= the man is sleeping).

Note that the words that are translated as English objects in the first and third sentences have the same case morphemes as do the subjects of the second and fourth sentences. We have seen, however, that these so-called objects are, in Hidatsa, the subjects of the verbs, and the apparent subjects of the first and third sentences are actually the subjects of the causitive element hē, which was deleted by (ii). Thus, the ergative, in itself, is not a fundamental relation in Hidatsa, but rather an automatic result of a fundamental distinction between two types of verb phrase and a general causitive construction. In Hidatsa, subjects are marked differentially by the kind of verb phrase that follows them, and objects, which may occur in both types of verb phrases, are marked the same, and only by position.

G. H. Matthews