STRUCTURE 0F THE GERMAN NOUN PHRASE

Problem

It has been pointed out many times that mechanical translation cannot be successful unless some attempt is made to find a procedure different from the mere replacement of words in the original text by words in the target language. It is true that one can obtain a sort of translation by using this method which amounts to little more than using an automatic dictionary, but such translations would require editing. It has been noted, therefore, that it will be necessary in developing a program for mechanical translation to make some provision not only for the replacement of words in the source text by words that are approximately equivalent in the target language but also for the replacement of syntactic structure in the original text by approximately equivalent structures. It will be necessary, therefore, to describe exactly the structures in both languages and to state which structures are equivalent, or approximately equivalent. The equivalences will be determined by comparing the distributions of similar structures in the two languages. Then it will be possible to formulate rules for converting the structures of one to the structures of the other. These rules will necessarily include rules for the arrangement of words within the structure and rules for placing the constituents of the structure in proper sequence in the larger environment of the sentence.

The languages with which we are currently concerned are English and German. Because both of those languages are Indo-European, it may not be so difficult to solve the problem of finding equivalent structures in them as it would be if they belonged to totally unrelated linguistic families. In

both languages it is possible to identify two major syntactic units: the noun phrase and the verb phrase. (This is true for the majority of sentences in the written language if not necessarily true for a large number of utterances in the written language).

The noun phrase is defined in this paper as a syntactic rather than a morphological unit. Morphologically, a noun phrase in German may be defined rather easily as a sequence of morphemes whose head is a word that can occur in capitalized form in non-initial position. Such sequences have a number of syntactic functions. They may occur before or after verbs as subjects or objects, after prepositions and in some cases after other nouns. The definition of a noun phrase as a syntactic unit is different, however. In this case we are concerned with a sequence of morphemes that functions as the subject of a finite verb. The morphemes or words in the sequence may belong to any one of several morphological classes. The only member of the noun phrase that may be said to be constant is the final affix of the finite verb. Using the rules of German word order makes it easier to locate the noun phrase, for it is either the first or third phrase of the sentence if the sentence is neither interrogative nor imperative*. Special rules must be formulated for interrogative sentences. For imperative sentences the noun phrase is the second phrase of the sentence. In the present study only those noun phrases whose final member is the verbal affix -t or one of its alternants will be considered, i.e., those that may be classified as third person singular.

---

* Examples: Dieser Wein ist besser.
Hier ist er.
Jetzt geht er nach Hause.
Der alte Mann gingØ langsam den Fluss entlang.

Procedure:

It is possible to construct a series of sentences that fall within the limits prescribed in the definition of the noun phrase. In addition to satisfying the conditions specified in the definition the sentences must be such that it is possible to alter the structure of each sentence by changing one structural feature at a time. These sentences will then serve as substitution frames for the various constituents of the noun phrase, The sentences, with the noun phrase completed in each case by aiding one of the elements to be tested, are presented to a native speaker of German who has been instructed to indicate which of the sequences can occur. The results give a picture of the distribution of the constituents in terms of the structures in which they can occur. The constituents are then classified according to their distribution. A further result of the analysis is the classification of structural features in a sentence in terms of the effect of a given feature on the occurrence of a class of noun phrase elements.

The words used in this test have been taken from A Grammar of the German Language by George O. Curme (2nd revised edition).

It should be made clear at this point that the results obtained still may not be considered complete even for the type of noun phrase defined. A further analysis will require the examination of a large body of texts. The results obtained so far can, however, serve as a basis for the formulation of tentative structural statements.

The words are divided into groups according to the description given in Curme as articles, determinatives, demonstratives, numerals and descriptive adjectives. Rules have been stated in traditional grammars for the uses of these forms, but in almost all cases, the rules have been concerned only with the immediate environment of the particular class of words rather than with the structure of the sentence as a whole.

Additional elements of the noun phrase still remain to be considered. Relative clauses, appositives and prepositional phrases are among these elements. At the end of the study, it will be possible to present a formal statement of form classes with exact rules for the combination of the members of the classes so that given a list of words and the rules, it will be possible to construct all permissible sequences which can be called noun phrases. Conversely, given a sequence in German, it will be possible to state exactly the type of structure that the sentence represents. This is important for mechanical translation because it shortens the time required for the recognition of the sentence structure before it can be matched with the equivalent English structure.

J.R. Applegate
Department of Modern Languages
and
Research Laboratory of Electronics
Massachusetts Institute of Technology