The C.M.T. (Code Matching Technique) Mechanical Translation Process

by

Ariadne Lukjanow

In mechanical translation the following components are crucial in the process of language transfer from source to target language:

1. the transfer of grammatical forms,
2. the transfer of meaning and
3. the rearrangement problem (syntax).

All of these components and the entire analysis has to center on the most important problem, the transfer process.

In arriving at a solution the grammatical categories of all languages involved have to be analyzed, and the synthesis of the analysis that will permit the transfer process will result in a common grammar that has all the pertinent problems of all languages after a process of reducing the synthesis.

The synthesis is carried out by analyzing each of the languages involved, that is of the individual parts of speech of the same.

Then a comparison is made of the results of the analysis in terms of the transfer problem; functional parts of speech are devised that are common to the two specific languages to be translated, as well as establishing other grammatical categories such as cases, tenses, genders, etc.

The synthesis is then a bilingual grammar where the normal nomenclature of the conventional grammar is being used, but where the components are not merely the identical components as in conventional grammar of any one specific language taken separately. It also contains new categories not existing in either language, and the old components may have an entirely different meaning.

In order to make this bilingual grammar available to the machine process, the categories are reduced to a code that can be accepted by a machine. From now on, a machine does not have to analyze the various forms,but it can merely use the code that already contains the answer to the problem of transfering grammatical categories.

At the same time as the categories have been devised their interrelationship has been established in order to permit an intelligent breaking up of the text to be translated into meaningful segments by grammatical boundaries.

In the process of meaning transfer one for one equivalents have to be determined for every possible situation and variation. This in turn leads to establishing certain classes of meaning pertaining to every part of speech of the above mentioned bilingual grammar. So far, slightly more than 300 of such classes have been established. The classes have been reduced to a three digit code and have been incorporated into the overall code. Those codes for the classes provide the specific meaning of any member by a purely numerical comparison in the process of translation, rather than try to attempt a meaning analysis by the machine.

In the rearrangement process a simple enumeration of the members of the sentence (words, etc.) has been fitted into a system which will permit the rearrangement problem to be carried out as a strictly numerical operation. The classes devised pertain strictly to the target language, but the application of those numbers-classes is a function of the source language.

In the process of grammar and meaning transfer redundant information may be supplied, and a specific lack of information unique to the target language may exist.

Prior to the actual rearrangement redundant information is deleted by a process of recognizing certain identical parts of the codes. The lacking information is supplied through a numerical analysis of the output codes so that under certain conditions the missing information in the source language such as is, was, they, them, it, etc., may be directly inserted in the proper place.

The actual matching process is carried out by numerical manipulation of the overall code. By nature this code supplies all of the possible answers, and the operation therefore is reduced to a numerical matching of the codes within the established boundries, where then identical pieces of codes are selected.

In this mechanical translation process the source language equals a code that is supplied through a collation against the dictionary. These codes are used in the matching process.

The output of the matching process contains the proper code which in turn through their dictionary line number elicit the proper target language translation by another collation against the target language dictionary. The output then results in the proper target language translation.