

MACHINE TRANSLATION IN REVIEW

Harry H. Josselson
Dept. of Slavic and Eastern Languages
Wayne State University
Detroit, Mich. 48202

"In order to attain the long-range research objectives of machine translation, it is important that research be continued on such short-range objectives as straightforward linguistic analysis of the lexicon and the grammar of various languages."

The primary aim of machine translation research is to produce the best possible translation, automated wherever feasible, from one language (the source language) into another (the target language) through the combined efforts of linguists, programmers, and research associates involved in related fields. A secondary aim, more an outgrowth of early research than an objective in itself, is to develop, as far as possible, a complete linguistic description of the grammars of certain languages. Accumulation of such data is invaluable for subsequent efforts to refine and develop machine translation output. Also, it is of great interest to linguists and teachers of the respective languages.

Historical Background

The idea of machine translation dates back to 1946, when Warren Weaver and A. D. Booth of the Massachusetts Institute of Technology began discussions on the technical feasibility of machine translation. A great deal of progress has been made since that time. In 1954, International Business Machines conducted a demonstration of machine translation in cooperation with the Georgetown University. An IBM 701 general purpose computer, using a total vocabulary of 250 words and six rules for determining the relationships that exist among sentence constituents, was programmed to translate Russian sentences into English. This highly publicized demonstration provided a marked impetus for interest growth and active research in the field. Machine translation research groups were organized not only in the United States, but also in countries throughout the world, e.g., Great Britain, France, Italy, Germany, the Soviet Union, and Japan.

Main Tasks

These pioneer researchers realized that, if any real progress were to be made in machine translation, linguists and programmers would have to devise a system to store and retrieve vast quantities of natural data in the form of dictionary entries and text, and to create algorithms to perform analysis

and synthesis. Also, they realized that high-speed hardware with vast memory capacity would have to be developed simultaneously with that system. The main tasks confronting the researchers, in ascending order of difficulty, are listed below:

- Compiling automatic dictionaries and developing procedures for efficient storage and retrieval of language data and translation rules or instructions (lexicon);
- Encoding the grammar of words and their constituent parts (morphology);
- Writing word order rules for sentence analysis (syntax or sentence structure determination or recognition); and
- Developing procedures for analyzing and codifying the meanings of words (semantics).

By 1960, about a dozen federally sponsored research groups in the United States were investigating problems connected with compiling automatic dictionaries, development storage and retrieval procedures, and morphology. Indeed, not until investigations were actually underway did the researchers realize just how little was known about the structure and usage of given languages or just how incomplete the descriptions of these languages were.

Machine Translation Conferences

In order to minimize duplication of tasks and to establish a basis for cooperative exchanges of information, a series of machine translation conferences was organized by Wayne State University. The first such conference was convened at the request and with the support of the Information Systems Branch, Office of Naval Research. Support for subsequent meetings was provided by the National Science Foundation, the United States Air Force and ONR.

The first machine translation conference, held in Princeton in 1960, concentrated on dictionary design and general questions of grammar. The second, which convened at Georgetown University in 1961, was devoted to problems of grammar coding. The third meeting, held in 1962 back at Princeton, was syntax-oriented; and the fourth, held in Las Vegas in 1965, dealt with semantics.

rules are revised in the near future to include provisions to rearrange word order and to insert articles and prewords. This latter task is still being done by a human post-editor. However, even in its present form, the telegram style output can be easily understood.

Experimentation in machine translation at Wayne State University is continuing with two primary objectives: first, to automate additional analytic procedures; and second, to refine previous routines according to insights gathered from each successive experiment. At the same time, translation rules are being formulated which, when written into the system, will improve the quality of the output.

Machine Translation Research Objectives

In order to attain the long-range research objectives of machine translation, it is important that research be continued on such short-range objectives as straightforward linguistic analysis of the lexicon and the grammar of various languages. Information to be gained from such research is necessary to successfully implement machine translation procedures and to overcome those problem areas or stumbling blocks that stand in the path of completely successful machine translation. These problem areas are primarily concerned with syntax (the structural interrelationships of sentence constituents) and semantics (the meaning of words and groups of words). There is reason to believe that sound theoretical bases have been established which will eventually resolve syntactic problems. However, a great deal remains to be done in semantics.

In any system of machine translation, there are many points of intersection in the areas of lexicon, morphology, syntax, and semantics. The boundaries are not clearcut but have had to be mapped and defined. The machine translation research process has been a heuristic one, where the knowledge and experience gained at a given stage of development has been applied to the refinement of preceding stages. The converse is also true. In this regard, the initial exploration of semantic problems, in many instances, has been based on syntactics.

Organizations Involved in Computer-Aided Translation

Although fully automatic, high-quality machine translation is still a remote objective, machine-aided translation (*i.e.*, translation produced through a symbiosis of man and mechanical devices) is indeed workable and productive. This is especially true in scientific areas for which microglossaries and microgrammers have been compiled.

To date, five organizations have been conducting effective computer-aided translation operations:

- Atomic Energy Commission at Oak Ridge, Tennessee;
- Foreign Technology Division, United States Air Force, Dayton, Ohio;
- National Physics Laboratory, Teddington, England;
- EURATOM at Brussels, Belgium and Ispra, Italy; and
- Central Research Institute for Patent Information, Moscow, USSR.

It is interesting to note that the last-named institute is translating the *Official Gazette*, the weekly publication of the United States Patent Bureau.

The operational machine-aided translation programs being conducted at the above organizations continue to produce a good deal of useful output. However, there is still a need for computer-aided linguistic investigations such as those being conducted at Wayne State University and at the Linguistic Research Center of the University of Texas. The results of these investigations are needed to refine further machine-translation-oriented linguistic analysis and to implement those refinements into operational programs.