SPECIAL SECTION ON
MACHINE TRANSLATION OF
NATURAL LANGUAGES

Sergei Nirenburg
Guest Editor

Introduction

The field of machine translation has recently entered a new, third period in its evolution. In its early period, for roughly fifteen years from 1950 MT was an expanding field of study in which both research and development efforts were undertaken. It is well-known and well documented (Bar Hillel, 1960; ALPAC, 1966) that this early MT paradigm could not and did not produce fully automated high quality translation systems. In fact, the practical results were quite negligible for such a high-scale effort. This period also witnessed an engineering-oriented trend which considered MT primarily an engineering problem. In the positive sense, it helped promote the idea of nonnumerical computation. As another consequence, the engineering, nontheoretical attitude to MT has been inherited by the next period in MT history. A more significant and lasting influence was exerted by the less "practical" aspect of early MT research: it helped achieve advances in theoretical and computational linguistics as well as artificial intelligence.

The widespread disappointment with the practical results of the early MT research ushered in the second period of MT, the post--ALPAC one. This second period of MT extended from roughly 1966 until recently. The philosophy of this period was to largely abandon the goal of fully automated high quality translation and generally to set and pursue only those goals that were immediately feasible. The grand design of the interlingua approach, so popular with earlier MT, failed to produce tangible results at the then current level of knowledge of linguistics, artificial intelligence and computer science. As a natural reaction to this, the methodological framework for all MT work in its second period has been the so-called transfer approach, which allowed a certain (rarely very deep) level of linguistic analysis of the input text and relied on the bilingual lexical and grammatical structure dictionaries to perform actual translation. This paradigm was accepted and polished in the well-established European MT centers in Grenoble and Saarbruecken and within a recent international EUROTRA project funded by Common Market and the METAL group at the University of Texas.

As a result of this shift of emphasis toward the feasible, MT got a reputation (at least in the U.S.) of a research topic not suited for the academy. People who chose research careers in MT in the 50s went to AI or computational linguistics instead in the 70s and 80s. (Additional information about the early and recent history of MT can be found in Hutchins, 1982; Buchmann, 1984; Slocum, 1984; Tucker & Nirenburg, 1984.)

Recent years witnessed a marked resurgence of interest in MT as both a practical engineering problem and, most importantly, as a research field within linguistics and artificial intelligence. Thus, EUROTRA's philosophy, although emphatically transfer-based is that of a "linguist's", not "engineer's" project (cf. King and Perschke, 1984); extensive application of linguistic and knowledge representation methods and ideas characterizes the massive Japanese effort in MT (cf. Nagao, 1982). Several modest attempts were made to study the problem of MT in the framework of artificial intelligence (Carbonell et al, 1981; Lytinen & Schank, 1982 and few others; cf. also Shann, 1984 for a discussion coming from a person outside AI and the perspective in Slocum, 1984). It would be fair to say, however, that the potential AI holds for this application has not yet been tested. At the same time, MT's promise as an application and a verification testbed for AI ideas has largely gone unnoticed.

There are many reasons, scientific, technological and sociological for MT to enter this third period of its development, where the concept of a language-independent intermediate representation of text, interlingua, may become much more feasible than before.

Significant advances have been made in the fields of theoretical and computational linguistics and artificial intelligence. On the computational side, formalisms and techniques for morphological, syntactic and semantic (conceptual) parsing of natural language have been extensively studied. The area of knowledge representation within AI has produced results important for the study of semantics and pragmatics of natural language Many important techniques of knowledge management have been suggested, and a number of them tested in relatively non-trivial applications. This is an incursion into the realm of encyclopaedic knowledge which Bar Hillel, in the 1960s, considered to be the main obstacle to the success of MT.

Computer science and technology have made spectacular progress in the areas of hardware, computer languages, databases, word and text processing and information systems, thus making even the "brute force" approaches far more feasible than they were twenty years ago.

A sociological motivation is expressed in that MT is among the objectives of Japan's Fifth Generation initiative and Europe's Common Market developments for the next decade. Several research groups are now working in this field in Japan, with ample funding and valuable results have already been produced. EEC's EUROTRA project also promises significant progress. Its characteristic feature is multilinguality (seven languages are involved, both as sources and as targets).

The purpose of this Special Section was not only to show what MT projects are active all over the world, but also to elicit opinions on how the long-term basic research projects in MT should be conducted and what lessons from linguistics and AI should be learned by MT practitioners. The concrete questions of this nature constituted the second part of the detailed questionnaire that was distributed among potential participants (the complete text of the questionnaire is given below). These questions will be discussed at some length during a conference on theoretical and methodological issues of machine translation that will be held at Colgate University in August 1985.

The questionnaire elicited 12 responses. Some of the MT research projects chose not to participate (EUROTRA, METAL). Some others may have stayed ignorant about this special section. Some workers may have thought that the time they must spend answering the questions in the second part of the questionnaire is more than they could afford. At any rate, the picture this section offers is not complete. Thus, for example, no reports are included about...
the state of MT in the USSR (one possible source is Marchuk, 1984). Also, the response rate of AI researchers has not been overwhelming. We hope that the information in this Section will attract the attention of the AI community to MT as a desirable application area.

One half of the responses come from active MT practitioners, the rest, from linguists. The reports are arranged alphabetically by the institution.

**Bibliography.**


**QUESTIONNAIRE**

for an entry in the SIGART Newsletter
Special Section on Machine Translation

**PART I. REPORT**

1. Title of project
2. Affiliation
3. Sponsorship
4. Date started
5. Personnel
6. Address(es)
7. Spokesperson (if applicable)
8. Status of project: operational; experimental; development; other
9. Aim: fully automatic MT, Machine-Aided Human Translation, Human-Aided MT, other?
10. Source language(s)
11. Target language(s)
12. Field(s) of translation
13. A 10-line abstract of the project content.
14. Design scheme: transfer, interlingua, direct, other?
15. What type(s) of internal representation of the syntax and/or meaning of the input text do you use? If possible, give examples.
16. For each stage of the translation process (morphological, syntactic, semantic, other types of analysis; transfer; generation) please answer the following, mentioning the linguistic and computational approaches developed independently of the project and modified for use in your system (and, of course, specifying the approaches developed within the project itself):
   a. what types of grammars are used and what language phenomena do they cover?
   b. what formalisms (description languages) are used for specifying the grammars?
   c. how is the computer implementation of this module designed?
17. What are the sizes and design peculiarities of the dictionaries used in your project?
18. How do you deal with non-Latin alphabets?
19. Implementation
   a. equipment (computer facilities, operating system)
   b. programming languages and environment(s)
   c. size of program(s).
20. Working Environment. Describe the auxiliary systems used (or developed and used) for increasing productivity, modularity and other parameters both during the development process and at the operational stage.
21. Performance information. What throughput has been achieved so far? What are the major bottlenecks? What are the prospects for future improvements? Other facts (if applicable), e.g., SL text input rate; pre- and post-editing rate; interactive editing rate; Operation costs, etc.
22. Bibliography.

**PART II. DESIDERATA FOR MT**

Please describe your position on the following questions:

1. MT as an Application Area of Theoretical Linguistics:
   a. what should the proper relationship between theory and applications be in Linguistics?
   b. can the current syntactic and semantic theories be utilized in designing an MT system and if so, then which or what parts of them?
   c. what procedures should one use for adapting the linguistic theories for use in an MT application?
   d. is there a connection between MT and other application areas of theoretical linguistics?
   e. is there a need for developing a linguistic theory of translation? If so, what main issues should this theory address?
   f. should we study linguistically the concrete pairs of source and target languages in order...