LANGUAGE AND MACHINES

COMPUTERS IN TRANSLATION AND LINGUISTICS

A Report by the

Automatic Language Processing Advisory Committee

Division of Behavioral Sciences

National Academy of Sciences

National Research Council

Publication 1416

National Academy of Sciences National Research Council

AUTOMATIC LANGUAGE PROCESSING ADVISORY COMMITTEE

John R. Pierce, Bell Telephone Laboratories, Chairman
John B. Carroll, Harvard University
Eric P. Hamp, University of Chicago*
David G. Hays, The RAND Corporation
Charles F. Hockett, Cornell University†
Anthony G. Oettinger, Harvard University
Alan Perlis, Carnegie Institute of Technology

STAFF

A. Hood Roberts, Executive Secretary
Mrs. Sandra Ferony, Secretary

*Appointed February 1965

Contents

August 20, 1965

Dear Dr. Seitz:

In April of 1964 you formed an Automatic Language Processing Advisory Committee at the request of Dr. Leland Haworth, Director of the National Science Foundation, to advise the Department of Defense, the Central Intelligence Agency, and the National Science Foundation on research and development in the general field of mechanical translation of foreign languages. We quickly found that you were correct in stating that there are many strongly held but often conflicting opinions about the promise of machine translation and about what the most fruitful steps are that should be taken now.

In order to reach reasonable conclusions and to offer sensible advice we have found it necessary to learn from experts in a wide variety of fields (their names are listed in Appendix 20). We have informed ourselves concerning the needs for translation, considered the evaluation of translations, and compared the capabilities of machines and human beings in translation and in other language processing functions.

We found that what we heard led us all to the same conclusions, and the report which we are submitting herewith states our common views and recommendations. We believe that these can form the basis for useful changes in the support of research aimed at an increased understanding of a vitally important phenomenon—language, and development aimed at improved human translation, with an appropriate use of machine aids.

We are sorry that other obligations made it necessary for Charles F. Hockett, one of the original members of the Committee, to resign before the writing of our report. He nonetheless made valuable contributions to our work, which we wish to acknowledge.

Sincerely yours,

J. R. Pierce, Chairman
Automatic Language Processing
Advisory Committee

Dr. Frederick Seitz, President
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

July 27, 1966

Dear Dr. Seitz:

In connection with the report of the Automatic Language Processing Advisory Committee, National Research Council, which was reviewed by the Committee on Science and Public Policy on March 13, John R. Pierce, the chairman, was asked to prepare a brief statement of the support needs for computational linguistics, as distinct from automatic language translation. This request was prompted by a fear that the committee report, read in isolation, might result in termination of research support for computational linguistics as well as in the recommended reduction of support aimed at relatively short-term goals in translation.

Dr. Pierce's recommendation states in part as follows:

The computer has opened up to linguists a host of challenges, partial insights, and potentialities. We believe these can be aptly compared with the challenges, problems, and insights of particle physics. Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.

There is every reason to believe that facing up to this challenge will ultimately lead to important contributions in many fields. A deeper knowledge of language could help:

1. To teach foreign languages more effectively.
2. To teach about the nature of language more effectively.
3. To use natural language more effectively in instruction and communication.
4. To enable us to engineer artificial languages for special purposes (e.g., pilot-to-control-tower languages).
5. To enable us to make meaningful psychological experiments in language use and in human communication and thought. Unless we know what language is we don't know what we must explain.
6. To use machines as aids in translation and in information retrieval.

However, the state of linguistics is such that excellent research that has value in itself is essential if linguistics is ultimately to make such contributions.

Such research must make use of computers. The data we must examine in order to find out about language is overwhelming both in quantity and in complexity. Computers give promise of helping us control the problems relating to the tremendous volume of data, and to a lesser extent the prob-

Therefore, among the important kinds of research that need to be done and should be supported are (1) basic developmental research in computer methods for handling language, as tools to help the linguistic scientist discover and state his generalizations, and as tools to help check proposed generalizations against data; and (2) developmental research in methods to allow linguistic scientists to use computers to state in detail the complex kinds of theories (for example, grammars and theories of meaning) they produce, so that the theories can be checked in detail.

The most reasonable government source of support for research in computational linguistics is the National Science Foundation. How much support is needed? Some of the work must be done on a rather large scale, since small-scale experiments and work with miniature models of language have proved seriously deceptive in the past, and one can come to grips with real problems only above a certain scale of grammar size, dictionary size, and available corpus.

We estimate that work on a reasonably large scale can be supported in one institution for $600 or $700 thousand a year. We believe that work on this scale would be justified at four or five centers. Thus, an annual expenditure of $2.5 to $3 million seems reasonable for research. This figure is not intended to include support of work aimed at immediate practical applications of one sort or another.

This recommendation, which I understand has the endorsement of Dr. Pierce's committee, was also sent out for comment to the membership of the Committee on Science and Public Policy. While the Committee on Science and Public Policy has not considered the recommended program in computational linguistics in competition with other National Science Foundation programs, we do believe that Dr. Pierce's statement should be brought to the attention of the National Science Foundation as information necessary to put the report of the Advisory Committee in proper perspective.

Sincerely yours,

Harvey Brooks, Chairman
Committee on Science and Public Policy

Dr. Frederick Seitz, President
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

In computational linguistics and automatic language translation, we are witnessing dramatic applications of computers to the advance of science and knowledge. In this report, the Automatic Language Processing Advisory Committee of the National Research Council describes the state of development of these applications. It has thus performed an invaluable service for the entire scientific community.

Frederick Seitz, President
National Academy of Sciences