Mechanical Translation Research
Statement of Victor H. Yngve,
Massachusetts Institute of Technology
before the House Committee on Science and Astronautics, May 11, 1960

Mr. Chairman and members of the committee, I appreciate the opportunity to be the first scientific investigator on your list of witnesses on the topic of mechanical translation. Since this position has fallen to my lot, I feel it would be most helpful if I reviewed for you the field as a whole and the extent of progress as I see it, as well as the way in which our own work fits into this broader picture. I shall proceed by trying to answer some questions that are frequently raised.

People often ask why we are interested in the eventual possibility of translating from one language to another by the aid of machines. It has, of course, great intrinsic scientific interest. Since a sizable portion of the funds being expended today in this area comes from military budgets, I assume that some people feel that there would be extensive military applications. Without appearing to contradict this view, and without denying the possibility of other aims within the military, I should like to state that in my view the eventual development of mechanical translation will have a far greater impact in its peace-time applications than it will have in military applications. May I expand on this briefly?

The world is divided by language barriers into about 4000 linguistic communities. Many of these communities are small and represent primitive or underdeveloped cultures. But well over 50 of these, language communities are large and important enough to carry on extensive trade, communication, and cultural interchange one with another. All interchange between language communities must now funnel through individuals who are to some extent bilingual. The resulting bottle-necks serve to stifle such intercourse and to keep the language communities in comparative isolation. It has been suggested that the adoption of a single universal second language - either a natural language or an artificial one - would offer a solution. I agree that this would be ideal. The various language areas could continue to maintain their linguistic integrity at the small expense of having to learn only one second language in order to communicate with the rest of the world. There have been hundreds of attempts to introduce such a universal second language, and they have all failed because it has been impossible to get enough people to agree on the idea or on the choice of a language. So we are left with the necessity for a great deal of translation in order to conduct the day-to-day business of normal peace-time activities. In this area, which includes scientific and technical communications as well as other kinds, the eventual possibility of mechanical translation could be a great boon.

Mechanical translation would involve the use of the automatic digital computer - either one of the existing general-purpose machines or eventually, when we know more about it, a special-purpose machine. The machine operations involved are allied to other types of information-processing by machine. They include access to dictionaries and tables stored in a large machine memory, and appropriate automatic manipulation and processing of words and items. The only trouble is that we don't yet know how to do it.

There has arisen in the past ten or twelve years a small band of pioneers dedicated to finding out how to instruct a machine to produce satisfactory translations. Support has been adequate. I know of no one with a sensible research proposal ever being for long without adequate support. But at the same time, I feel that the supporting agencies have been sufficiently careful in screening applications so that there has been very little if any waste of government funds on ill-conceived projects.

Work in mechanical translation can be separated into three parts: science, technology, and production. Under science we have research directed toward the discovery of the basic facts and knowledge of languages and translation that will form a firm foundation for erecting a technology. Under technology we can include research leading to the development of the dictionaries and machines that our science tells us how to build. Under production we would of course contemplate actual use of the technology for the production of useful translations.

Taking these three areas of endeavor in reverse order, I would like to say a few things about each. First, on production. I know of no system of mechanical translation, either existing or proposed, that would be capable of yielding adequate translations in the next few years. By adequate translations, I mean translations competitive with those made by humans. There are, of course, a number of systems under development. They all have serious defects, as their proprietors would be the first to admit. The government should be exceedingly cautious in assessing any scheme of mechanical translation that is alleged to be ready for production. The reason is that it is very difficult to assess accurately the merit of a less-than-perfect translation system. May I repeat, I know of no system that would be capable of yielding adequate translations in the next few years.

Second, in the area of development, there are a number of groups throughout the country working very hard. A lot of good work is being done. These groups are building a technology. They are putting dictionary information into computers; they are building machines. This work is important. They are experimenting with sentence analysis and playing with semantics. Their motto is "Let's do now what we

know how to do. Maybe someone will be able to use it." Maybe. But I suggest extreme
caution because the results at present are not good enough. If someone wants to use
one of these systems, let him do it with his eyes open, counting the costs and
counting the error rate.

Prospective customers are likely to say that a less-than-perfect product will
be satisfactory. Perhaps a product with 90 per cent or 95 per cent accuracy. We
hear such figures. Any quoted per cent accuracy means very little because of the
difficulty of assessing less-than-perfect output. The trouble is that even if errors
could be counted, it is difficult to determine the relative amount of loss caused by
different kinds of errors. Some errors are not serious, others are very serious.
But if I were pressed for a figure I would say that realistically we can't reach an
accuracy of 50 per cent at present. But even if we could achieve 95 per cent accuracy,
what would it mean? Would it mean that we would miss the 5 per cent of important
new material and get the 95 per cent of already known material? There is some
indication that this would be the case.

I don't want to be misunderstood. I am not advocating perfection where perfection
is perhaps unattainable. But would we tolerate a mechanized bank accounting system
that guaranteed 95 per cent accuracy in crediting and debiting the accounts? Would
we tolerate voting machines that guaranteed a 95 per cent accuracy in totaling the
vote? Would we tolerate a mechanized post-office that would guarantee correct
delivery of 95 per cent of the letters? the 5 per cent of error might be caused
by the failure of simple rules-of-thumb for treating the multiple meaning of words
like "Washington." Such a rule-of-thumb might be: "Send a letter with "Washington"
in the address to Washington D. C. if the letter is mailed east of the Mississippi,
and to the state of Washington if mailed west of the Mississippi, and ignore such
infrequent meanings as Washington, Georgia; Washington, Illinois; Washington,
Indiana; Washington, Iowa; Washington, Missouri; Washington, New Jersey; Washington,
North Carolina; and Washington, Pennsylvania, since the error rate will scarcely be
affected." It is such rules for dealing with the multiple meaning of words that
lead to unreliable translations. If someone can use such a system, I will not
object, but let him look it over carefully first.

The groups that are developing systems are performing a vital service. They
are building the technology that we will surely need, but it must be on as firm a
scientific foundation as can be found. The trouble is that there is no foundation
at all in certain places. This brings us to the third area of mechanical translation
research, science. The group at M. I. T. and some of the other groups are working

very hard trying to build an adequate foundation. We are working on the science of communication. Our motto is "Let us isolate those areas where our routines are imperfect for lack of basic knowledge, and go after that basic knowledge." We are seeking to fill in the gaps and to expand and deepen our understanding. We are seeking routines that are intrinsically capable of producing correct and accurate translations. (This will be a long, difficult process, but a necessary on if satisfactory mechanical translation is ever to be a reality.

At M. I. T., research on mechanical translation is conceived of as just one facet in a rapidly emerging area of study - the Communication Sciences. Pioneering work is being done in a number of other areas that fit under the Communication Sciences. Work is being done in Artificial Intelligence, Communication Biophysics, Communication Systems, Experimental Psychology, Linguistics, Neurophysiology, Processing and Transmission of Information, Sensory Aids for the Handicapped, Social Science, and Speech Communications. Advances in any of these related fields are quite likely to be directly or indirectly applicable to mechanical translation. Needless to say, the presence at M. I. T. of a number of scientists contributing in the related areas of the Communication Sciences provides an ideal environment for the conduct of mechanical translation research.

Our mechanical translation research has been sponsored primarily by the National Science Foundation. Our relations with this agency have convinced me that the Congress was very wise in setting up the Foundation in the way that it did. They have been able to attract an excellent staff. All of the people that I have dealt with at the Foundation have proved to be exceptionally competent and enlightened public servants. While doing their job well, they have interfered in no way with the scientific conduct of our research. This very enlightened attitude on their part has helped greatly in our being able to maintain at M. I. T. an atmosphere favorable to the individual creativity so necessary for the nurturing of basic scientific advances.

We have made a number of advances. The list of areas where we have made significant advances is long and is a matter of record. I will not bore you with a complete recitation; it is available if you want it. I would like to comment, however, on a few items.

The work on generative grammar and the theory of grammatical transformations by A. N. Chomsky represents an important advance in linguistics. It provides an approach to syntax that sheds considerable light on some questions of meaning. It sets a new standard of scientific rigor and precision in the description of languages

to replace the old approximative and normative type of grammar.

Recent work by E. S. Klima, of my group, in his continuing study of English grammar points out formal features common to many diverse expressions that are negative in meaning. These formal features are themselves best accounted for by assuming a certain family of grammatical categories that correlates not only with the general notion of negation but also with that of varying degrees of complete-ness of negation. The formal features extend from such words as "not," "none;" and "never" through such prefixes as "un-" and "dis-" to words commonly characterized as inherently negative like "doubt" and "forbid." A correct understanding and treatment of negation will be essential for correct translation, and for a proper understanding of how our language works.

A number of other important topics have been worked on or are under continuing investigation. Practical problems in computer research have not been ignored. Work on the COMIT system in cooperation with the M. I. T. Computation Center makes available a powerful programming language for use in linguistic work. By the use of the computer as a research tool, linguists will be able to do many things that have hitherto been impossible.

I should like to conclude by discussing some of my own work. I have recently discovered a possible explanation for many previously puzzling complications in sentence structure. It is now possible to explain why English has a part-of-speech system; why it has both active and passive; why it is that when a verb has two objects, a direct and an indirect object, the one that is a pronoun comes first; why we use the anticipatory "it" as in "It is true that he went;" why a relative clause follows its noun while adjectives precede; why it is that of the two genitive markers, the one that precedes is a separate word (of), and the one that follows is a suffix ('s); why it is that awkward sentences are awkward. The explanation is quite simple and is now being published in the Journal of the American; Philosophical Society.

Only the surface has been scratched. There are many unanswered questions to be investigated: What further generalizations can be made in syntax? What is meaning, and how do we encode and decode messages in a language like English? For what purposes is language used, and how does the use of language affect translation? What is the nature of the translation relation and how can it be described or specified? To what extent are languages translatable? By what methodology can we achieve des-criptions of languages? What are the detailed facts making up the descriptions of the various languages? The answers to these questions will be important not only for the development of accurate and trustworthy mechanical translation, but they will provide new insights into the mechanisms of language and thought processes, the very foundation of human culture.