DOCUMENT RESUME

ED 359 942 IR 016 164

TITLE	Machine Translation Technology: A Potential Key to the Information Age. Report of the FCCSET Committee on Industry and Technology.
INSTITUTION	Office of Science and Technology Policy, Washington, DC.
REPORT NO	PB-93-134336
PUB DATE	Jan 93
NOTE	61p.
AVAILABLE FROM	National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.
PUB TYPE	Reference Materials - Bibliographies (131) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Annotated Bibliographies; *Automation; *Computer Oriented Programs; Computer Software; Editing; Foreign Countries; *Government Role; *Machine Translation; Policy Formation; Productivity; Scientific and Technical Information; *Second Languages; *Technological Advancement; Word Processing
IDENTIFIERS	Japan; *Natural Language Processing; United States

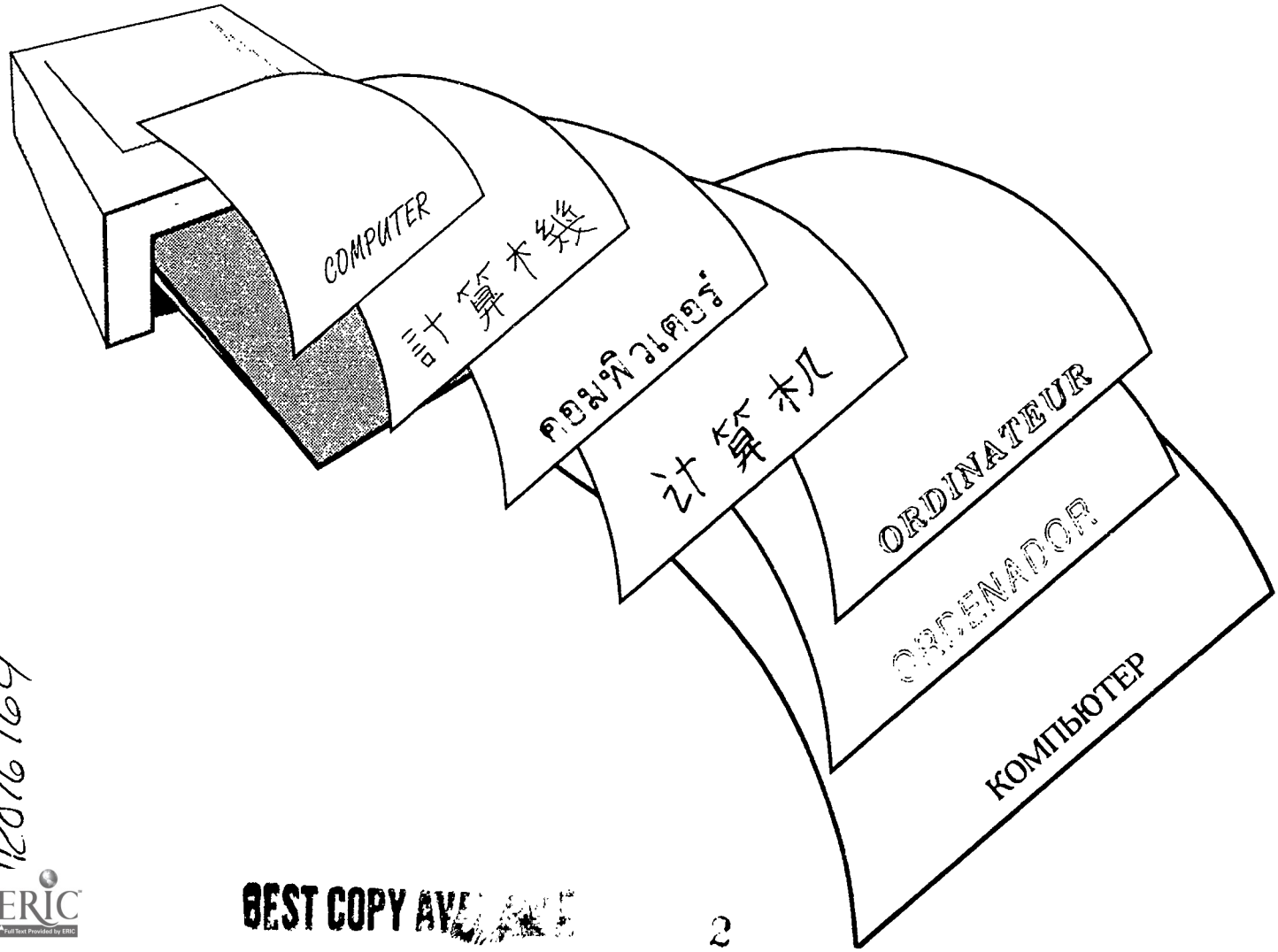
ABSTRACT

Machine translation (MT), an emerging technology that enables text to be translated from one language to another by computer, represents an indispensable contribution to the sharing of technical information particularly since nearly half of the world's scientific and technological literature is written in languages other than English. The state of the art and its potential are discussed. No existing MT system appears capable of producing polished translations without some human involvement, but current systems can yield definite benefits in improved productivity in certain situations. The United States is strong in research on natural language processing, but faces the challenge of converting its research potential and knowledge into commercial operating systems. Japan, where 14 commercially viable systems have been developed, is far ahead in this area. The U.S. Federal Government has a demonstrable need for foreign-language information and should consider devising polices and strategies to become a world leader in MT. The government could be a catalyst for MT research by encouraging the involvement of U.S. industries in developing and commercializing products and services. Three figures illustrate the discussion. Appendix A lists members of a working group on MT. Appendix B lists 77 annotated selected sources on MT. (Contains 20 references.) (SLD)

Machine Translation Technology

A Potential Key to the Information Age

FCCSET Committee on Industry and Technology



Machine Translation Technology

A Potential Key to the Information Age

Report of the FCCSET Committee on Industry and Technology

January 1993

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20506

I am pleased to release "Machine Translation Technology: A Potential Key to the Information Age," a report by the Committee on Industry and Technology (CIT) of the Federal Coordinating Council for Science, Engineering and Technology (FCCSET).

The Information Age requires us to understand ideas and documents produced beyond our nation's borders. Leaders in both the private and public sectors in the U.S. need to have access to foreign-language information in a form they can comprehend in order to successfully meet the challenges of global competition. Machine translation, the emerging technology that enables text to be translated from one language to another by computer, represents an indispensable contribution to the sharing of technical information.

Today nearly half of the world's scientific and technological literature is written in languages other than English. As advances are made, it is essential that all members of the science and technology community be aware of and understand them. Human translation is scarce, expensive and cannot keep pace with the current explosion of information around the world. Machine translation can help to solve this problem for scientists, as well as for experts in other fields.

Currently, machine translation can produce comprehendible translated text in a few languages, but human editing is often required to produce a truly polished translation. In limited applications, machine translation has been valuable both in its capacity to determine if a text need be perfectly translated for future use and to significantly simplify the work of translators. Machine translation can streamline access to foreign-language information for our domestic purposes, and it can allow materials published in the United States to be understood throughout the world. The facts given in this report establish machine translation as a vital information technology.

Extensive effort was necessary on the part of senior government officials, private experts and researchers to collect and integrate information from many sources for this report. The CIT Chairman, Dr. Robert M. White, and the interagency membership of the Committee and its Working Group on Machine Translation have done an excellent job and are to be commended.

D. Allan Bromley
Director

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20506

FOR IMMEDIATE RELEASE **JANUARY 15, 1993**

For more information contact:
Joseph Clark, Department of Commerce, 202-482-4844
Elizabeth Rodriguez, Office of Science and Technology Policy, 202-395-5101

MACHINE TRANSLATION TECHNOLOGY WILL AID U. S. COMPETITIVENESS

The report, "Machine Translation Technology: A Potential Key to the Information Age," released today by Dr. D. Allan Bromley, Assistant to the President for Science and Technology, describes Machine Translation (MT), which is a language translation generated by a computer, with or without human intervention. MT can provide fundamental assistance in meeting the challenges of today's global Information Age and can have a significant impact for the U. S. and international competitiveness.

According to Dr. Bromley, "Research and development on machine translation will advance the nation's position in artificial intelligence, natural language processing and related fields." Dr. Bromley noted, "This report discusses the importance of MT technology and how important it is for the U. S. to advance and augment its use."

Highlights of the report include advances that the U. S. will gain both nationally and internationally from pursuing MT technology:

- **Today, nearly 50% of science and technology literature is being published in foreign languages. With MT, scientists as well as experts in many fields writing in different languages will improve their ability to share information in collaborative international efforts.**

- **Some 70% of information in a given technology is found in patents and 8 million patents filed abroad are not in English. With the aid of improved MT technology, the United States could gain immediate access to a vast amount of brand new technology.**

- **For the purpose of exporting American products to non-English-speaking countries, business correspondence, technical and repair manuals, and advertising materials could be prepared into the importer's language more efficiently and consistently with the aid of Machine Translation.**

- MT is a new high-tech industry which will create high-skill jobs in the United States.
- Advancing MT technology will augment developments in both computer hardware and software.
- Related technologies, including artificial intelligence, optical character recognition, and natural language processing will experience innovation boosts with the advancement of MT.

The report was developed by the Working Group on Machine Translation of the Committee on Industry and Technology which is part of the Federal Coordinating Council for Science, Engineering and Technology. Participating agencies include the U.S. Departments of Commerce, Defense, Energy, Interior, State and the National Science Foundation.

Copies of the Machine Translation Technology report may be obtained by the public from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161. Phone 703-487-4650. Ask for Report #PB93-134336.

TABLE OF CONTENTS

Executive Summary

Information on which leaders can take action is critical to resolving the challenges of our global Information Age. But business and government leaders cannot act on foreign-language information they cannot read or understand. Nearly half of the world's scientific and technological literature is now published in foreign languages.

Human translation is adequate for occasional translation needs. But it is quite slow, expensive, and often inaccurate when a high volume of foreign language material overwhelms editorial control. Something like machine translation (MT)— defined as "translation generated by a computer, with or without human intervention"— can help. Modern computer technology can potentially play a major role in moving information across language barriers.

No extant MT system appears capable of producing polished translations without some human involvement. But current systems can yield definite benefits in improved productivity in certain situations. For example, a raw translation quickly produced by MT can be given directly to a consumer for determining whether further translation is necessary. Moreover, terminology banks, grammar checkers, and bilingual word processing software can provide leverage for human translators.

MT may become a vital enabling technology for the Information Age. Part of its potential impact lies in its promise for advancing related technologies, such as natural language processing and optical character recognition.

U.S. Strengths and Weaknesses. The United States is strong in basic research on natural language processing, and leads the world in the diversity of language pairs under development. The challenge is in converting its formidable research potential and know-how into commercial operating systems. The single most challenging competitor-nation in MT is Japan. Japan is ahead of the United States in development of MT systems— in terms of the number of systems, variety of applications, and development of knowledge sources— but not in basic research. Japan also leads the U.S. in commercial use and general acceptance of MT, and in integration of MT into the office environment.

There is a gap between America's need for translation and the resources it devotes to it. In the last 15 years, the Japanese government spent over $200 million to help develop commercial MT products. In contrast, each year the U.S. Government spends only a few million dollars on MT development— a fraction of what Japan or Europe spend.

The U.S. Government has a demonstrably increasing need for foreign-language information. This report highlights the major known translation needs and activities of intelligence, defense, and civilian agencies, but it is not comprehensive. These descriptions of translation activities reflect only part of the real demand. Information specialists report that users frequently cancel their requests for translations when the work cannot be done (by human translators) in a timely fashion.

U.S. applications of MT are from foreign languages into English (to assimilate information) and MT from English into foreign languages (to disseminate information). From foreign languages into English, there is a monitoring function to avoid technological surprise, both for strategic defense and for protection of U.S. markets. America's need for MT from English into foreign languages is largely to support U.S. industry in marketing products overseas. MT could augment U.S. efforts to increase sales abroad and thus enable companies to be more competitive.

Development of MT Systems. MT research began in Europe and flourished in the 1950s and early 1960s in the United States. After Sputnik in 1957, MT projects began at some 20 U.S. institutions. But computer technology in the 1960s was judged inadequate to overcome the linguistic complexities, so broad Government support virtually disappeared. However, some intelligence-agency support of MT research continued after 1966, and all three MT systems that today dominate the Western Hemisphere market—SYSTRAN, LOGOS, and METAL— survived.

Worldwide, 25 companies are selling MT systems that are known to be installed and in regular use; 7 are recognized to have U.S. roots. In Europe, the MT market is dominated by SYSTRAN and METAL, both developed in the United States and modified in Europe. Seven years ago, the Commission of the European Communities launched the largest single MT undertaking ever, the $30 million Eurotra project. Eurotra's aim was to provide rapid translation among all member nations' languages. The project met with many frustrations and is being dismantled.

Although research in Japan on MT began around 1956, it did not burgeon until the 1980s, fueled by the dynamic growth in Japan's trade with the West. The inaccessibility of the Japanese language to foreigners, and the need for a more efficient way to process the volume of scientific and technical information available in English, led the Japanese to view MT as key to Japan's growth. There are 25 Japanese organizations known to be developers of MT systems; 14 of these systems are commercially viable.

U.S. Government Role. The pace of global developments in technology and trade means that the United States should soon consider devising policies and strategies for MT in order to retain world leadership. If the Government decided to promote the development of MT technology, one role it could play would be as a catalyst. Perhaps through coordinated investment in MT as a critical information technology, the Government could stimulate increased research on MT and encourage the involvement of U.S. industries in developing and commercializing products and services.

Conquering the barriers to using foreign-language information will help to strengthen American presence in worldwide developments. In that effort, MT technology would make a significant difference.

Introduction

The world is at the threshold of an Information Age. Actionable information is critical to resolving today's global challenges- not only strategic military and economic concerns, but also questions of housing, education, environment, and health. But the United States is laboring under an explosion of information, much of it from abroad. For instance, nearly 50 percent of the world's scientific and technological literature is now published in foreign languages: Japanese, Russian, German, French, Chinese, Korean, and Arabic, with Czech, Hungarian, Polish, and other East European languages soon to join the list.

Real-time access to and evaluation of foreign scientific, technical, military, and economic information can help to shape U.S. national and industrial policies. But traditional human translation cannot keep pace with this information explosion because it is slow, expensive, and sometimes inaccurate (especially when translating new technical terminologies). Nor is all of this information really useful— translating it all is not only impossible but wasteful. It needs to be screened before investments are made in full-text translation. Machine translation (MT) can be an excellent scanning tool for this purpose. Modern computer technology has the potential to play a major role in moving information across language barriers.

No extant MT system in the world is yet capable of producing polished translations without human involvement. But current systems can yield definite benefits in improved productivity in specific situations. For example, a raw translation quickly produced by MT can be given directly to a consumer for determining whether further translation is necessary. Moreover, terminology banks, grammar checkers, and bilingual word processing software can provide leverage for human translators.

Research and development (R&D) in machine translation is attractive for many reasons. It represents "good science," which is now at a crucial juncture for potential breakthroughs. It is also a locus of possible international cooperation in advancing the state-of-the-art in MT itself, as well as in several generic information technologies, such as natural language processing, artificial intelligence, and optical character recognition. As a critical information technology, MT could spin off benefits for society in such areas as health, science and technology (S&T), and the environment.

As an *enabling technology*, MT could open new vistas for the S&T community to exploit. In research, MT technology could advance the frontiers of scientific knowledge in such areas as artificial intelligence, knowledge processing, and computational linguistics. MT and its associated technologies center around information processing. Translation is often a crucial element of information processing and all of the facets of human language processing that go with it, including optical character recognition, speech input and output, and intelligent human interfaces. MT development provides what researchers consider the ideal test bed for these technologies. MT can contribute to the development of more powerful natural language software, not only across language barriers but also for English alone; it provides insights into the linguistics of such technologies as information scanning, abstracting, and free-text searching.

MT could enable the production of innovative information products and services to cross language barriers. Science and education, too, could profit through better assimilation and dissemination of knowledge, and through improved tools for acquiring and using languages. National defense and security could be aided by faster and more thorough monitoring of foreign information. For industries, MT could strengthen those that exist and create others that are information and knowledge-based.

MT currently provides a small window for efficiently capturing and using state-of-the-art information from a few other nations. The United States could best explore what lies beyond that window through the development of MT that meets broader national needs. These needs may be different in emphasis from those of Japan. Japan's impetus for MT development was generated by that country's requirement to translate into customers' languages documentation that accompanies exports. U.S. exporters share this need, but America needs MT even more to assimilate information from abroad to sift through voluminous, fast-changing material in many languages and speedily produce crude translations for American users, so these users can determine what might be significant for them. Without MT that suits America's needs, the Nation could lag in its awareness of foreign technology. Many observers believe that, for most foreign languages, particularly Japanese, the demand imposed by the information explosion far outstrips the available cadre of human translators.

The pace of worldwide developments means that the United States should soon consider devising policies and strategies for MT if it is to retain a world leadership role. Indicators abound of heightened activity abroad. For example, the U.S. share of international patents for new discoveries will soon drop below 50 percent, a signal that in the future much significant S&T information will have to be gleaned from foreign, not English, language sources. Japanese patents often contain cutting-edge information on technological advances, unavailable elsewhere. An MT system could expedite access to the 300,000 patent applications filed annually in Japan, and would allow U.S. companies to protect their own innovations from encroachment by seeing whether foreign companies are unfairly appropriating U.S. intellectual property rights. It could also help the United States acquire information from the growing number of Japanese databases.

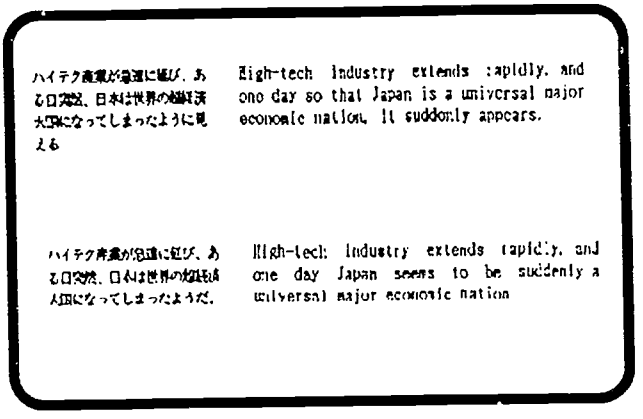
A prime example of foreign investment in MT development comes from Japan. In the last 15 years, the Japanese government spent over $200 million to help develop commercial MT products that combine computer hardware and software. In addition, 16 Japanese companies including such major firms as Fujitsu, Hitachi, Toshiba, Oki, NEC, Matsushita, Mitsubishi, and Sharp are investing in MT. Their systems will help the Japanese economy progress by translating foreign information into Japanese and Japanese into foreign-language instruction manuals for exported products. In contrast, the U.S. Government has measured its annual expenditures on MT only in hundreds of thousands of dollars (see footnote)— a very small proportion of the funding levels of either Japan or Europe. Some Government MT experts have expressed the opinion that what is needed to advance MT technology at the Government level are more tangible contributions to existing projects, with steady funding and participation by more agencies.

Clearly, there is a marked disparity between America's need for translation in all its forms, and the resources it makes available to conduct and pay for it. The Federal Government could consider playing a catalytic role in targeting MT as a critical generic technology for cross-cutting investment, and in promoting the involvement of U.S. industries in its development. One of the principal arguments for a policy for Government support of MT development is the need to filter the mass of foreign information and channel the right material to the right people quickly. The importance of this "brokering" function cannot be understated. In short, a strategic investment by the Federal Government in planning, coordinating, and supporting MT development will greatly contribute to America's competitiveness in the global Information Age.

Description of MT and Related Technologies

Translation in Perspective. The task of translation is to convey an idea and its associated meanings from source language to target language. Translation involves more than linguistic analysis and disambiguation. It is also an exercise in cultural transfer, animated by human intuition and instincts— traits that, without human post-editing, are lacking in today's MT systems. Yet the humor underlying the phrase "something lost in translation" pertains as much to human translation as to machine translation. Claims to reading or speaking a language do not necessarily mean that one can *translate*. Like simultaneous interpretation, translation is a special skill in which relatively few excel. Even among professional language officers, the skill level usually ranges only from good to very good. The shortage of translation skills in such languages as Japanese and Arabic, the problems human translators face in keeping current on technical terminologies, and the evolving information explosion, all underscore the need for computers to help overcome language barriers.

MT is not designed to replace human translators. Instead, it could fill a critical niche in a natural language processing environment providing the enabling technology for quick, accurate information while dovetailing with the traditional skills of human translation.



Example of machine translation. Adapted from the JTEC Machine Translation Workshop, Washington, D.C., March 8,1991

What Is Machine Translation? Machine translation is generally defined as "translation generated by a computer, with or without human intervention." It can be combined with various forms of human intervention at any point in the translation process. In some cases, it is used raw, with no human intervention whatsoever. The distinguishing characteristic of MT is that the translation starts out as an electronic file generated automatically by a computer. For this to happen, the text to be translated (the source text) must be submitted to the computer in machine-readable form. In human translation, the target text is generated by a person, who needs no input devices to read the source text.

In the early days of MT, it was assumed that fully automatic high-quality machine translation (FAHQMT) could ultimately be achieved if enough time and energy were spent on building computer-based lexicons, or dictionaries, and sets of linguistic rules. Although quality did improve over the years, the analysis of language by computers began to reveal far more complexity than had ever been conceived, and FAHQMT became an increasingly distant concept. High expectations gradually shifted to a recognition that, for many purposes, human-assisted MT was the most that could be hoped for.

Human-assisted MT can take several forms. One is pre-editing, in which either the text is revised to eliminate problems known to baffle the computer, or a new text is written in a customized way. In a second form, post-editing, the machine's raw output is polished by a translator, editor, or technical expert. A third form is interactive editing, in which the operator is queried on-line and responds while the machine does its work. Whether people are used to enhance the computer's performance depends somewhat on the quality of raw output from the computer, but a very important factor is how the translation will be used. When the translation is for assimilation,i.e. to monitor information, standards can often be relaxed. When the translation is for dissemination and must stand extensive scrutiny, then the human element is called for.

Human translation can also make use of computer tools, in which case it is called machine-assisted human translation. This term usually denotes use of on-line computer systems. It can also include telecommunications, desktop publishing, spelling checkers, grammar checkers, and so-called machine pre-translation but it does not include any software that actually generates a text.

Basic Approaches to MT Development. In terms of linguistic design, there are basically three types of MT systems: direct, transfer, and pivot (or interlingua). These designations reflect the linguistic philosophy behind the systems' development. They parallel the evolution of scientific insights into language.

In the early years of MT (officially, MT history began in 1933, even before the existence of electronic computers), the systems were *direct*, translating mainly word-for-word. The rules that shaped the translation were part of the program itself and were invoked at random, wherever they happened to "work." An example was the Russian-English system developed at the University of Washington in the 1950s, which failed when it was tried at the U.S. Air Force's Foreign Technology Division. More recently, there has been a spate of PC software

that also fits this description. Because these systems have no basis in scientific linguistics, they do not allow for the fact that languages have an underlying structure that needs to be captured in order to produce translation. Direct systems never progressed very far, and are considered to have few advocates. Some that started out being direct, such as SYSTRAN and SPANAM, quickly evolved into transfer systems.

The *transfer* system is so called because there is a shared interface, or intermediate representation, that bridges the transition from source language to target language. Thus, it is a three-stage process: (1) source-language analysis, (2) transfer into an intermediate representation, and (3) target-language generation. This design makes it possible to plug in different target languages, and in some cases different source languages as well. The latter possibility depends on the robustness of the transfer component, the kinds and amount of syntactic and semantic information it covers, how much work it does, and the range of languages it can handle. Transfer systems today cover many possibilities. Almost all the commercial MT systems use the transfer model, which by now has a good record for both domain-specific and general-purpose MT.

Philosophically, the transfer system is a blend of the two basic linguistic approaches to MT development— the theoretical and the empirical. For over 30 years, debate has focused on the merits of these approaches. In the beginning, the MT scene mirrored the tensions prevailing in scientific linguistics in the late 1950s and early 1960s. The theoretical approach was advocated by followers of Noam Chomsky (1957), who advocated introspection about the nature of language. Their point was that language is not the linear expression that we hear or see in print, but rather the deeper structure behind it, which obeys its own intrinsically ordered rules— a structure which they believed to be universal. On the other hand, the empiricists, citing the known advantages of working with naturally occurring data, sought to formulate linguistic principles based mainly on evidence afforded by real text. Eventually, the more extreme positions were tempered and it became clear that both perspectives are essential to create an MT system that works.

The *pivot* (or interlingua) design, which for the last six years has been challenging the widely entrenched transfer design, shifts the emphasis back toward the more theoretical. Its key is a central component containing universal linguistic rules and semantic knowledge which, performing bidirectionally, translates from any language into any other language. In practical terms, the pivot design offloads much of the information normally contained in the source and target modules and concentrates these data in a shared knowledge base. The idea is theoretically attractive, and there are convincing practical arguments in its favor as well. It makes good economic sense in situations requiring translation from many languages into many (for example, the European Commission's nine official tongues yield 72 possible combinations). This was the rationale behind the large Eurotra project, which is now winding down. It is also the basis for R&D at Carnegie Mellon University, New Mexico State University, MCC, Fujitsu, and Japan's Center of the International Cooperation for Computerization (CICC). The pivot approach is especially suited for translation tasks in limited domains, in which the capacity to draw on world knowledge will enable the MT system to generate reliable output that should require little post-editing.

Is MT Crucial to U.S. National Interests ?

Vision. Without MT, a country unable to assimilate a high volume of potentially useful information from abroad could lack timely, accurate data and lose its edge in international business, diplomacy, military readiness, and academic research. Thus, the United States could benefit from MT capabilities and related technologies that can systematically filter raw information and produce rough translations of significant developments. For example, the DARPA Tipster Project extracts data in the original language then provides an MT-translated output. MT is considered the ultimate enabling technology for the Information Age, because it provides the ideal test bed for the parsers that underlie all natural language processing.

MT's maximum potential impact results from its crucial role in the information industry. Many observers believe that an organization positioning itself to become a leader in information processing must have MT. Experts argue that it is critical that the United States not give up as a key player in that industry, for several reasons. First, whoever controls the information industry will decide who has access to what. Second, the industry promises to be lucrative and to produce jobs. Third, since software drives hardware sales, if the United States abandons its development of MT software, it will lose even more of the hardware market than has already been lost. In contrast, the Japanese industrial giants are working concurrently on both MT and massively parallel supercomputers; they may be able to put these two technologies together and thus control both markets.

Recent technical assessments indicate that Japan is ahead of the United States in development of MT systems, but not in the research ideas. In fact, according to experts, some Japanese systems represent old technology and have encountered a plateau of effectiveness. The United States has the know-how but needs resources to convert it into operating systems. Most Western MT enterprises have been so strapped for resources that they, too, have very old technology. But the United States could "leapfrog" past the rest of the world and build MT systems that can use both multi-sentence contextual information and domain knowledge to overcome the limitations of the earlier transfer-based systems.

The challenge in developing MT will be to embed it into a system that will (1) extract information from unstructured texts for systematic storage in databases, and (2) classify the processed information for instant dissemination to U.S. consumers. When fully operational in the 21st century, such technology could monitor trends in national security, health, terrorism, world stock markets, gold fixes, and weather without human intervention.

Differing Applications. A clear distinction should be made between MT from foreign languages into English (to *assimilate* information) and MT from English into foreign languages (to *disseminate* information); the goals, priorities, gains, and tradeoffs are entirely different. From foreign languages into English, there is a monitoring function to avoid technological surprise, both for strategic defense and for protection of U.S. markets. For example, information contained in fresh patents for which the shelf life is only about six months may be critical to maintaining a competitive edge.

In many ways, MT could be ideal for keeping up with foreign technical literature. Human translators are apt to be expensive, slow, and scarce. Librarians and information specialists in government and industry consistently report that requests for translations are frequently withdrawn when they cannot be processed in a timely manner. These specialists readily admit that such requests would increase if prompt translations could be obtained. In contrast, MT automatically generates large volumes of *raw text* at a fraction of the cost of human translation, and at very high speeds. Indeed, MT may sometimes be the only way to screen massive amounts of material while it is still useful.

Currently, 63 U.S. Government agencies monitor foreign technical information, and several industries have systematic programs for keeping abreast of developments abroad. Their use of MT systems for crude translations could not only speed the delivery of information but also reduce duplication of effort and cost. Both basic linguistic research and long-term intensive development are needed to build the large, "try-anything" systems necessary for this task, especially for Japanese, Arabic, Chinese, Farsi, and Korean.

America's need for MT from English into foreign languages is largely to support U.S. industry in marketing products overseas. Aside from linguistic questions, in many cases the systems are specialized and domain-specific, and the basic (non-customized) technology is already well advanced. Further improvements will be incremental and largely application-specific. But the output quality needs to be better for dissemination than for assimilation, and current systems usually require much pre- and/or post-editing to improve quality. Better technology, not simply less expensive machines, could address this problem. What is holding back large-scale implementation is not so much the need for linguistic development as the fact that potential buyers are balking at the high cost of mainframe MT and are waiting until better, user-friendly systems come out on PCs. They also recognize that there is much training involved: users need to be taught how to update the dictionaries, troubleshoot the system, and make efficient changes in raw MT output.

Commercially, MT could augment U.S. efforts to increase overseas sales. It could enable companies to be more competitive in (1) providing technical documentation in customers' languages, and (2) speeding products to international markets by reducing the long delays of human translation. Businesses may also view MT as an efficient means of solving the problems of multi-language correspondence with foreign offices, although the technology cannot yet produce good translations in this area. The Japanese have successfully used MT for some years to document and promote their products in other languages.

In the United States, machine translation of documents is usually in a specific subject area and is from English into several foreign languages. Companies have built extensive dictionaries to support the specialized terminology, and they post-edit and sometimes pre-edit the documents. Potential buyers of MT systems are concerned about the cost of retraining translators to handle these new editing requirements. To avoid the expense of mainframes, many are waiting for less-expensive PC versions or are sending documents to MT service bureaus. Significant trade-offs in such choices involve cost, computational power, user interaction and user requirements.

According to some experts, the United States will never have MT that meets its needs unless it assumes responsibility for tailoring the technology to those needs. In other words, the Nation must "get in on the ground floor" in developing modern MT; otherwise, there will only be MT technology suited for European or Japanese needs, which can be somewhat different. Every MT application is for a different purpose, and in each case the system needs to be either custom-built or customized. The *customization* is the technology. Progress in MT development occurs through the processing of millions of words of text in the targeted application. Thus, the United States could be in a strong position to meet its specific requirements if it continues to support its own "home-grown" MT development, thereby controlling the construction of dictionaries and user-friendly interfaces.

MT is not important simply as a stand-alone "black box." It also has important uses as an embedded component of a wide array of information processing systems. Some experts maintain that it is usually difficult to take black box packages, without source code, and use them in all the ways that one wants. For instance, Japanese vendors are not giving users access to their code now; there is no reason to expect this to change. Thus, if Americans want to be prepared for a whole spectrum of possible, even unanticipated uses for MT, they need systems they own and can manipulate as necessary.

Americans' need for MT may differ in emphasis from that of other countries. The U.S. needs MT more strongly (but not exclusively) for scanning literature published in foreign languages. To do this, the United States would need to develop general-purpose systems capable of handling free syntax in a broad range of subject areas vastly more complex than anything the Japanese must have to market their products abroad (or can readily develop, given the shortage in Japan of native English speakers). Working alone, the Japanese cannot possibly "take over" MT technology to the extent of producing what America needs. Experts thus argue that the United States must either cooperate with Japan in this effort, or develop the technology by itself. Otherwise, U.S. needs will never be met.

Role of MT in Commercial Development

MT as a Commercial Opportunity. MT is already recognized for its contribution to the launching of U.S. products in multiple overseas markets. This is the case of English-into-many, which, as just noted, is not a priority for investment in innovative linguistic research. At present, U.S. industry depends on human translation to keep abreast of innovations abroad. Foreign-into-English MT could make a significant contribution in this area, but not without major linguistic development, especially from non-European source languages.

MT as an Enabling Technology. Researchers in natural language processing recognize that MT is the ideal test bed for the development of parsers, which resolve sentences into component parts of speech and describe them grammatically. Developments in MT contribute to speech recognition, free text searching, database abstracting, report-generating, and related areas of information management. MT can yield some spinoffs with commercial potential, including grammar checkers, report generators, and shells for developing specialized applications. Increased use of MT will increase the demand for accurate input text, thus increasing demand for improved optical character recognition, a technology increasingly employed to input typed pages into an MT system.

Evolution of Commercial MT in the United States. After a 1966 report by the Automatic Language Processing Advisory Committee (ALPAC) of the National Academy of Sciences/ National Research Council, which saw little merit in pursuing MT, public-sector support for practical MT in the United States evaporated and the private sector took over. For the first time, the technology was developed from the start as a commercial venture. The installation of SYSTRAN, the world's first entirely commercial system, at the U.S. Air Force in 1969 marked the beginning of a new era.

The other commercial pioneer was LOGOS, which began its activities in 1969. After developing its systems from English into Vietnamese and Farsi, both of them short-lived for political reasons, the company finally found a more lasting market with European languages.

The SMART Translator, the first MT system to be bundled with semiautomatic pre-editing software, reached the market in 1977. Weidner, founded in 1977, installed its first system in 1979. ALP Systems (now ALPNET) entered the picture in 1982 with the first and so far the only interactive MT system on the commercial market. Weidner made MT history when it launched the first PC system in 1983; GTS (formerly called Globalink) followed a few years later. ALPNET stopped selling its software in 1988, although it supports its original customers and uses its interactive MT, along with other translation-support software, in its chain of translation service bureaus. Weidner, bought earlier by the Japanese firm Bravice, was closed in 1988. Rights to the software were bought by InterGraph, and some of the language pairs have been updated and embedded in the company's desktop publishing platform. TOVNA, an Israeli-owned product launched in 1987 and sold in the United States by Translation Technologies International (TTI), claims to "learn" from examples provided by the user. The latest company to have a product on the U.S. market is Executive Communication

Systems (ECS), which recently announced its "MT Toolkit"-a shell with which users can develop customized systems. Other companies, such as IBM (and, in Japan, CSK), have developed MT systems for internal use.

Current Non-Japanese Vendors. Outside Japan, 11 companies are currently selling MT systems that are known to be installed and in regular use; seven of these are recognized to have U.S. roots. The table below lists the 11 firms in order of their appearance on the market:

COMMERCIAL MT SYSTEMS KNOWN TO BE IN REGULAR USE OUTSIDE JAPAN

Company	Product	PC	Origin	Ownership
Systran/Latsec	SYSTRAN	No	USA	France/USA/Japan
Logos	LOGOS	No	USA	USA
Smart	SMART Transl.	No	USA	USA
Chandioux	METEO 2, etc.	Yes	Canada	Canada
GTS	TWP	Yes	USA	USA
Atamiri	ATAMIRI	No	Bolivia	Bolivia
Siemens	METAL	No	USA	Germany
TTI	TOVNA	No	Israel	USA/Israel
Socatra	XLT	Yes	Canada	Canada
InterGraph[1]	DP/Translator	No	USA	USA
ECS	MT Toolkit	No	USA	USA

[1]Former Weidner software, PC versions of which are still used by original customers.

Inexpensive PC products for MT include: PC TRANSLATOR, TOLTRAN, SPANISH/ FRENCH/ GERMAN/ ITALIAN ASSISTANT, SPANISH EXPRESS, TRANSLATE, and PERSONAL TRANSLATOR. These products perform automatic word lookup; there is little or no morphological or syntactic analysis. They are less sophisticated than even the so-called "direct" MT systems that ran on mainframe computers in the early 1960s. Nevertheless they have been found useful by thousands of PC users.

Current Use. SYSTRAN is used by three large customers and a number of smaller ones. Each of the large users has a different type of application: the Air Force's Foreign Aerospace Science and Technology Center (FASTC) does information scanning of foreign texts; the Commission of the European Communities produces general translation of a variety of subject fields and document types, based on either "rapid" or "full" post-editing; and Xerox uses SYSTRAN to translate product manuals. LOGOS performs general-purpose translation for customers that include a translation bureau in the United States and the Canadian government, and it is also used for product manuals. Applications of the SMART Translator are largely (85 percent) for product manuals, and the company's front-end pre-editor is also sold separately to process input for other MT systems such as SYSTRAN. TTI's TOVNA is used for general purposes at the World Bank.

All six of the U.S.-owned companies in the table above have at least one installed customer base in the United States. All have customers abroad in Canada and/or Europe. The relatively low level of MT use in the United States is a reflection of the translation market in general, which not only is smaller than in the other "MT countries" but also involves fewer specialized text types in which the vocabulary and language structure are sufficiently predictable to be appropriate for narrow MT applications. This situation is changing, however, with the expansion of overseas markets for U.S. products, as well as the internationalization of both the supply and demand for translation. In the meantime, the use of general-purpose MT in the United States is confined to the Government and a few other sites, mostly translation bureaus.

Language Combinations. SYSTRAN has 28 language combinations either in use or under development at sites in Europe and North America: the five at FASTC are from foreign languages (including Japanese) into English, whereas Xerox has been working out of English into five foreign languages and will soon add five more. The EC has 16 language combinations, with English and other source languages, in different stages of development and use.

LOGOS initially concentrated on German and English (both directions) and then added other European languages in various combinations. Smart, GTS, Atamiri, and InterGraph have systems from English into various European languages, and GTS also has French and Spanish into English. TTI has English into French, with other languages under development. ECS has English and Korean (both directions) and English into Norwegian.

Philosophical Approaches. All of the commercial systems listed are transfer-based. So is Fujitsu's ATLAS II, soon to be introduced on the U.S. market.

SYSTRAN and LOGOS are both empirical in their approach, which means that they have the flexibility to expand their intermediate representations to include fully comprehensive linguistic information and knowledge bases. In both cases, the long history of practical use has produced context-sensitive lexical rule bases applicable to several language combinations, thus placing these systems in a position to handle a broad range of input texts Since its earliest years, LOGOS has relied heavily on a database of semantic information. The "MT Toolkit" developed by ECS is based on lexical-functional/unification grammar and is highly flexible within that paradigm.

Agenda for the Future. In all cases, the MT industry and users in the United States would benefit from enhancements to the systems' quality and flexibility.

Quality could be upgraded by expanding the scope and capabilities of the dictionaries. It could also be improved by building up the massive stock of rules linguistic and/or knowledge-based that dictate the choices made, whether these rules are embedded in the dictionaries or stored in separate tables. The need for dictionary-building is especially acute for general-purpose, "try-anything" MT. Currently, there is considerable duplication of effort, and activities are not necessarily in line with long-term priorities. In the future, increasing emphasis will have to be placed on coordination that leads to negotiated dictionary ex-

changes, availability of twin-text corpora in domains of greatest interest, and infusions of funding for key languages, text types, and subject areas where work lags behind.

Flexibility would be critical. Experts believe that MT must be poised for integration into desktop publishing, electronic services such as e-mail and on-line database retrieval, information scanning systems, and other computer networks. To keep up with this trend, some MT systems may have to be ported to standard platforms. Moreover, as the base of MT applications widens, a fundamental issue will be ease of customization. With the increasing involvement of users, especially non-linguists, systems will need more transparent interfaces. One of the biggest challenges will be to enable users to update the dictionaries easily and quickly, while invoking rich sources of linguistic information and world knowledge. At the same time, small, inexpensive systems can be developed for narrow applications; toolkits, if they are sufficiently user-friendly, will permit users to create their own software.

Among the options the MT industry might consider are expanding the technology's uses (e.g., to provide a correct dictionary lookup, complete with proper tense); developing larger dictionaries; sharing dictionaries; training translators in how to work with MT text; raising expectations for increased and faster translation; lowering expectations of quality for some kinds of translated technical text; increasing the number of language pairs; and enhancing accuracy.

Multinational Trends. The growth of worldwide marketing in general has been parallelled by the internationalization of translation services. This is particularly true of MT services, which rely on customization to be effective. Many of the MT companies have development sites in foreign countries, with responsibilities that range from dictionary-building to full-scale system development.

Major MT users often can build their own dictionaries. Sometimes, as part of this work, they acquire the capacity to contribute to other aspects of system development as well. This was the case, for example, with SYSTRAN at the EC, which ultimately became an autonomous developer.

Companies that sell their hardware overseas often establish local centers for preparing manuals. One such firm is Wang, which uses ATAMIRI in six different countries and, at some of these sites, has developed new language combinations from scratch. Although IBM has yet to introduce an MT product on the market, it has development sites in Japan, Israel, Spain, and the United States, and uses its English-Japanese system SHALT internally to translate hardware manuals.

In a slight variation of this trend, the MT vendors themselves have set up overseas laboratories to develop general-purpose systems in different language pairs. Siemens has followed this practice, with development sites for METAL in Belgium, Spain, Denmark, Germany, and the United States. LOGOS has development activities in Canada.

U.S. Government's Need for Foreign-Language Information

The following description of the U.S. Government's translation needs and activities is based on information and data provided in 1991. This description is intended to be illustrative, not comprehensive or exhaustive. These needs may represent only the tip of the iceberg in terms of the Nation's real demand for translations. The flow of foreign literature now includes important languages that professional technical translators in the U.S. are largely unable to translate (and those who can are expensive): Arabic, Farsi, Chinese, and Korean.

In general, among the Government's top priorities for language translation are increased versatility and flexibility, more language pairs, and large, constantly updated technical dictionaries in many disciplines.

Patent and Trademark Office (USPTO). On average, some 70 percent of the information in a given technology can be found primarily in patents. The rest is found in more traditional literature, such as technical publications and scientific papers. The USPTO, under the Department of Commerce, is developing an Automated Patent System (APS) to provide rapid information on U.S. and foreign patents. Unfortunately, over 8 million foreign patents filed abroad are not available in English; many of them must be translated to determine the patentability of foreign applications filed at the USPTO. With APS, a single database of all U.S. and foreign patents would give Americans access to crucial technical information at the touch of a button. Yet without high-speed automated translation, this system's potential could be limited, especially given the fact that the competitive importance of new patent information is highest during the first six months.

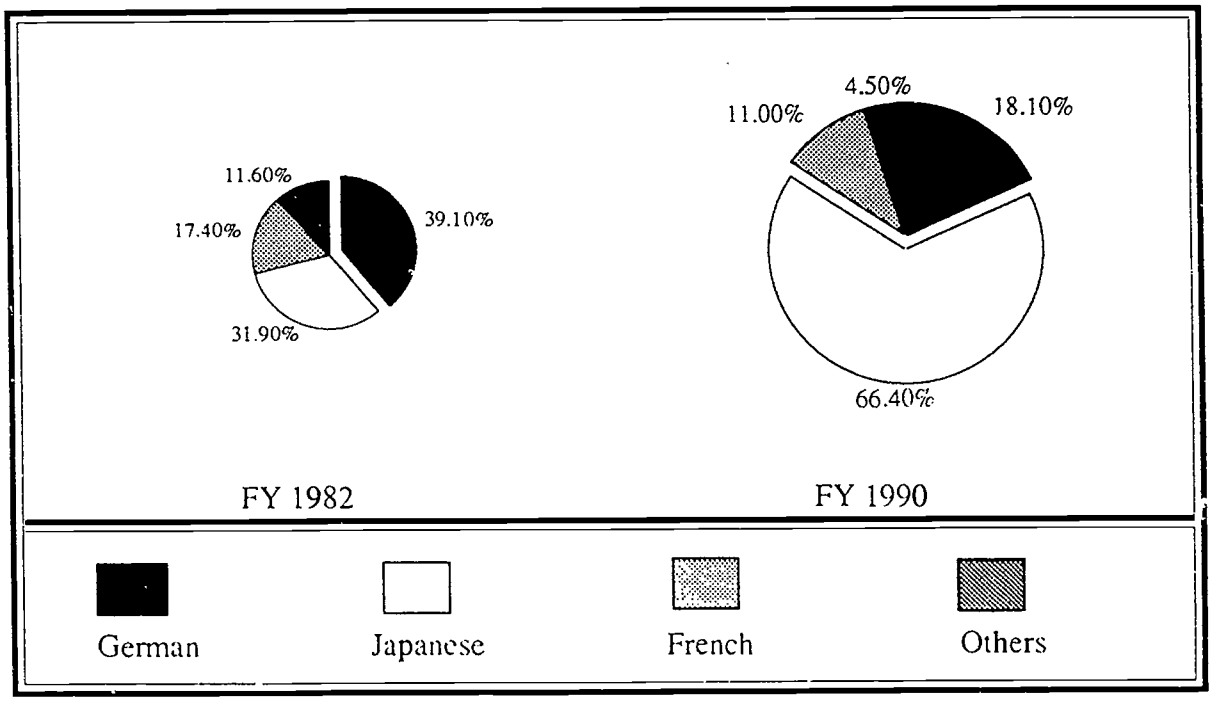
Foreign patent-holders, particularly Japanese applicants, benefit from the U.S. system's strong emphasis on protecting patentees' rights. Because of potent profitable markets, intellectual property law has recently burgeoned and become international. This trend is evident in 1989 USPTO figures: of the 102,692 U.S. patents granted, 46.7 percent were issued to inventors living outside the United States. Significantly, organizations receiving the most USPTO patents were Japanese corporations (Hitachi, Toshiba, Canon, and Fuji Photo Film); 21 percent of the U.S. patents were issued to residents of Japan.

The activities of the USPTO's Translation Section must be viewed against this background. They focus on translating foreign patents, written in foreign languages, that reflect R&D conducted overseas. To approve or decline a foreign application for a U.S. patent, USPTO examiners must check abroad to assess the status of the patent there, if it already exists, or the application's patentability, if new. As recently as 1980, the USPTO's translation services were made available to the public, patent attorneys in particular. In FY 1982, 1,572,504 words were translated: 39.1 percent German; 31.9 percent Japanese; 17.4 percent French; 4.7 percent Russian; 2.5 percent Italian; 1.3 percent Swedish; 1.2 percent Dutch; and the rest other languages. One translation company was contracted to supplement the in-house staff of five translators.

Since then, the Translation Branch's workload has skyrocketed. In FY 1990, the unit faced a written volume of 9,948,079 words (more than *six times* the 1982 demand): 66.4 percent Japanese; 18.1 percent German; 11.0 percent French; 1.6 percent Russian; 0.6 percent Italian; 0.5 percent Dutch; and the rest other languages. Available resources now include six translators (three solely for Japanese) and four contract translation firms.

U.S. Patent and Trademark Office Translations

Source: USPTO



Technology Administration (TA), Department of Commerce. The TA oversees the Japan Technology Program and the National Technical Information Service. Through these organizations, it is concerned with the development and availability of MT technology to increase access by U.S. researchers and strategic planners to translations of foreign documents, especially Japanese S&T literature.

Japan Technology Program (JTP). The JTP is part of Commerce's Technology Administration. In December 1989, the JTP provided $70,000 to fund a Japanese-English MT symposium organized by the National Research Council of the National Academy of Sciences. The symposium aimed to inform the U.S. audience about Japan's advances in Japanese-to-English MT, to share information and to stimulate thinking about how MT and related technologies might address U.S. users' needs. Professor Makoto Nagao of Kyoto University was the keynote speaker. Nagao is an internationally recognized expert and a pioneer in Japanese MT research, whose work has been supported by Japan's Science and Technology Agency and others. The symposium featured four panels, focusing on the state-of-the-art, market opportunities, users' needs, and R&D policy options. U.S. interest in MT may well have been catalyzed by the symposium.

In November 1990, with the National Science Foundation and Department of Defense agencies, the JTP sponsored a visit to Japan by an expert team to investigate Japanese MT activities and explore possible avenues of bilateral cooperation to advance the technology. The visit was coordinated under contract by the Japan Technology Evaluation Center (JTEC) of Loyola College, Baltimore, at a cost of $120,000.

National Technical Information Service (NTIS). NTIS is part of Commerce's Technology Administration. Of the documents NTIS acquires each year, 30 percent (23,000 reports) are from foreign sources; about one-third of those (10 percent of the total acquisitions) are in a language other than English. The agency performs only limited translation, because the cost must be amortized over a very small number of copies sold. An MT capability at NTIS could significantly expand the availability of foreign-language S&T information to U.S. users.

NTIS has cooperated with the Air Force Foreign Aerospace Science and Technology Center (FASTC) using SYSTRAN to translate Russian biotechnology information. Results of tests on a sample of journal articles were good. NTIS will have access to SYSTRAN through the Gateway System at the Defense Technical Information Center. French, German, and Russian language translation capability will be available. NTIS, along with the Japan Technology Program, has been active on the U.S.-Japan Task Force on Scientific and Technical Information, under the 1988 U.S.-Japan Agreement on Science and Technology. In 1991, NTIS began distributing a report on the JTEC team's findings from its MT survey visit to Japan in 1990.

As the Information Age progresses, NTIS may be naturally positioned to take the lead in addressing several important issues. First, with the expanded flow of information, how can the user community be encouraged to tap it most productively ? Second, how can rough translations of foreign information be produced quickly for users to scan ? And third, what are the most efficient and effective ways to channel specific information to the people who should most appropriately see it?

Defense Technical Information Center (DTIC). DTIC is the Department of Defense's central collector and distributor of all technical literature that results from DOD-funded activities in R&D and acquisitions. Also, DTIC has Memoranda of Understanding with some NATO countries and Sweden to exchange S&T literature. Between 7 and 10 percent of DTIC's collection (2,000 to 3,000 documents per year) comes from foreign sources. While some have been translated, many documents arrive with no English titles, text, or abstracts to indicate the content. Rough translations of such documents, generated by MT, would fill a void. Users would have quicker, more accurate access to foreign S&T information. DGIS will include the SYSTRAN MT system as a prototype for an on-line translation service; DTIC plans to install an interface to connect with Systran services at the Foreign Aerospace Science and Technology Center and with Latsec (the U.S. SYSTRAN developer) services in San Diego.

Defense Information Analysis Centers (IACs). DTIC manages 12 IACs, which review S&T literature worldwide. While information may appear in foreign languages, the IACs

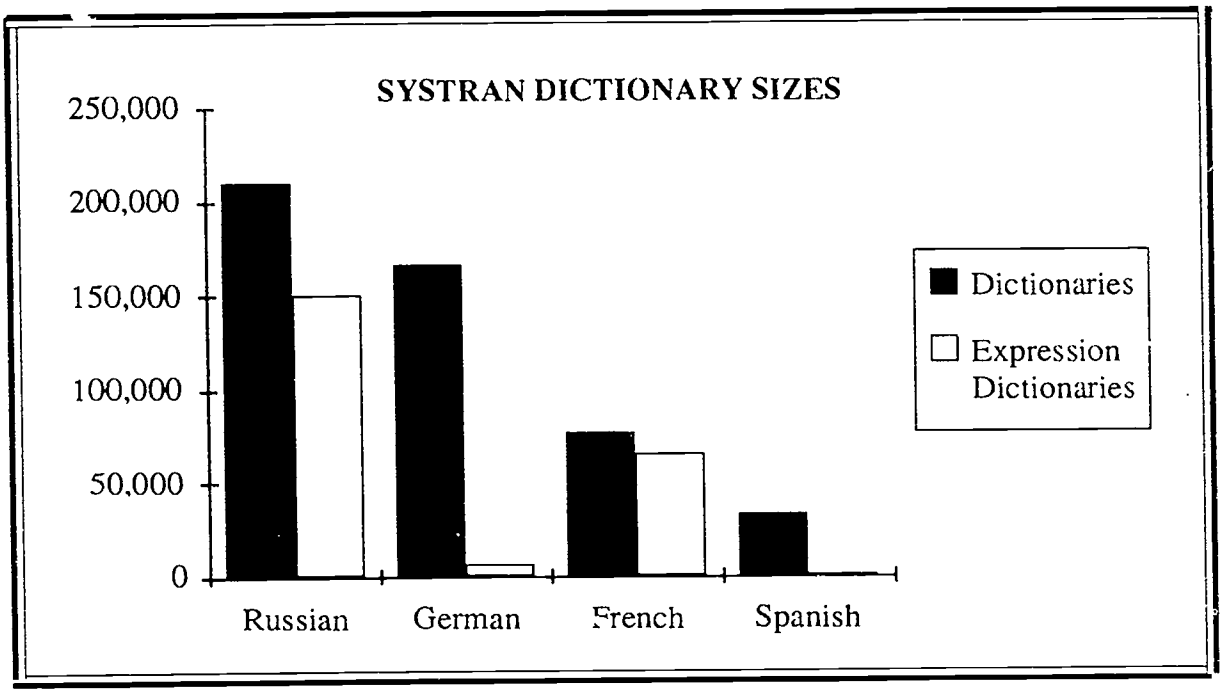
often cannot obtain translations. S&T materials in Russian, German, Japanese, and French need to be translated to track S&T development. The Foreign Science and Technology Center in Charlottesville, Virginia, provides translations of mutual interest to itself and the IACs.

The chief limitation in IAC use of foreign S&T information is funding for translations. IACs use contract translators on a minor scale, but have few or no funds for translating foreign-language materials. Recent budget restrictions make funding for translations even less likely. This situation reduces the scope of information the IACs can collect and evaluate. Lower translation costs with MT would increase the IACs' use of foreign information and help them achieve one of their objectives, i.e. to review S&T information worldwide.

Air Force Foreign Aerospace Science and Technology Center (FASTC). Located at Wright-Patterson Air Force Base, Ohio, FASTC maintains the Central Information and Reference Control (CIRC) databases of translated foreign information. FASTC monitors S&T developments abroad (traditionally in the USSR and Eastern Europe) and uses the CIRC databases to organize and distribute millions of records. Most of CIRC's foreign S&T information is machine-translated. All military and civilian intelligence agencies use CIRC as a primary source of foreign S&T information.

Over 20 years ago, DOD chartered FASTC to develop and operate an MT system, SYSTRAN (through a contract with Latsec), to translate S&T documents from Russian into English. The Russian-English pair has been installed and sponsored since 1969; sponsorship of French and German began in 1983 and of Japanese in 1985. A low level of Spanish-English sponsorship began in 1990. SYSTRAN now translates about 80 percent of FASTC's foreign language documents. SYSTRAN's FASTC version can translate Russian, German, and French quite successfully. SYSTRAN also offers these three language pairs commercially, as well as English into French, German, Spanish, Italian, Portuguese, and Arabic. SYSTRAN's English-Dutch pair is used by Xerox and the Commission of the European Communities. Other languages under development include Japanese and Korean, as well as English into Greek and Scandinavian languages.

The Russian-English system translates 50,000 to 60,000 pages of Russian text a year, in a process called focused information scanning. In this process, the MT system itself is used to screen volumes of text and zero in on the material to be translated. With the help of an automated post-editing program which highlights segments of the translation that need to be checked, only about 20 percent of the machine-translated text is reviewed by human translators, and an even smaller percentage is actually post-edited. Thus, researchers gain rapid access to technically accurate translations in many disciplines. Four years ago, even faster access to S&T information was ensured when raw SYSTRAN translation was made directly available to users at 1,400 terminals within FASTC. This on-line usage of SYSTRAN averages 185 accesses per month for Russian-English alone.



In December 1990, FASTC signed a development contract with Latsec to fund a five-year effort at $3 million. These funds include $300,000 per year contributed by the Foreign Broadcast Information Service (FBIS) to complete development of a Japanese-English system. The system is jointly supported by FBIS and the Intelligence Community Staff through FASTC.

Unlike most Japanese commercial systems that are based on workstation minicomputers, the SYSTRAN system works on a mainframe and is unique in its networking capability and speed. The current Japanese language system has a dictionary of about 50,000 words and expressions. It will first be used to translate titles and tables of contents of FASTC and FBIS texts.

As with the U.S. Government and industry overall, FASTC is shifting its emphasis to developing an ability to assess and react quickly to global situations. Although FASTC will continue to use MT to translate and report on long-term technological developments abroad, it is increasingly interested in using MT to gain quick access to current information on foreign technologies. For example, FASTC has been translating many Russian and German technical manuals acquired as a result of German reunification. During Operation Desert Storm, SYSTRAN was used to translate Russian and French materials, when the information had to be translated quickly. Since no organization had an Arabic-English MT system, any Arabic materials had to be sent to outside contractors. By the time the translations were completed, the war was over.

The project is continuously expanding the stem and expression dictionaries and refining the linguistic analysis. Current dictionaries for Russian, German, French, and Spanish contain 210,000, 166,000, 77,000 and 33,000 entries, respectively. Expression dictionary sizes for these languages are 150,000, 6,000, 65,000, and 500, respectively.

Foreign Broadcast Information Service. In FY 1990, FBIS translated roughly 63 million words from virtually all languages into English at a cost of over $3 million. MT was used to translate nearly half a million of these words from Russian and Japanese source material.

Military S&T Intelligence Centers. These Centers also translate foreign technical information or bibliographic references to it. The Consolidated Translation Survey (CTS) at the Foreign Broadcast Information Service indexes all translations by U.S. Government organizations. This index is available on-line through the Central Information and Reference Control (CIRC) database at the Foreign Aerospace Science and Technology Center.

Foreign Science and Technology Center. Military intelligence centers that use foreign S&T information from this Charlottesville center include the Armed Forces Medical Intelligence Center; Army Missile and Space Intelligence Center; Navy Technical Intelligence Center; Army Intelligence Agency and Army LABCOM; Los Alamos National Laboratory, International Technology Division; National Security Agency; and Defense Intelligence Agency.

Office of Under Secretary of Defense for Research and Engineering (OUSDR&E). In the early 1980s, the OUSDR&E sponsored a Task Force on Industry-to-Industry International Armaments Cooperation. The group targeted a need for a national cooperative program to translate key Japanese S&T and related policy documents, and proposed an expansion of U.S.-Japan cooperation in basic research. That DOD initiative has been superseded by the Department of Commerce's Japanese Technical Literature Program and the 1988 Cooperative Research and Development Agreement with Japan, under which advancing MT technology has been a stated priority.

In February 1986, the OUSDR&E produced an internal report, "Japanese Military Technology: Procedures for Transfers to the United States." It described mechanisms for gaining military technologies and information from Japan's government and industry. Japan has allowed the export of its military technology since 1983. MT capability could facilitate this effort.

Defense Technology Security Administration (DTSA). DTSA reviews export licenses for commodities controlled by export regulations. It needs information on international capabilities for producing strategic technologies on the Militarily Controlled Technologies List (MCTL). DTSA receives licenses from the Department of State for exporting munitions to other countries, and it reviews licenses for "dual-use" (civilian and military) technology processed by the Department of Commerce. There are cases where DTSA must perform substantial research on the quantity and quality of finished technological products from abroad. Faster translation of foreign technical information could accelerate the export-licensing process and enhance the U.S. position in international markets.

National Aeronautics and Space Administration (NASA). NASA's Scientific and Technical Information (STI) Program compiles and publishes bibliographies of aerospace informa-

tion from Japan, the Soviet Union, and Europe. Each bibliography focuses on aerospace S&T items that appear in NASA's public databases.

The STI Program contracts about 3 million words per year of translated S&T information. The international program acquires foreign S&T information by means of bilateral, trilateral, and direct agreements with foreign organizations and countries. The STI Program has exchange agreements with over 43 countries at the institutional level, and has third-party agreements with the United Kingdom, Sweden, Belgium, the Netherlands, and Italy through the European Space Agency (ESA). It also has direct, negotiated agreements with Australia, Canada, and Israel, and is negotiating similar agreements with Japan, India, and Hungary. It exchanges tapes, hard copy, and microfiche documents under these agreements and receives about 5,000 reports per year from ESA. This rate of foreign report acquisition is 25 percent higher than three years ago. The STI Program collects and enters these journal articles, conference proceedings, and technical government reports into the NASA/RECON aerospace database.

Engineers and scientists at NASA R&D Centers access this foreign information in the NASA/RECON aerospace database on a regular, usually time-dependent, basis. If a translation cannot be completed within the allotted time, requestors tend to cancel the actions, rather than pay for information that would be delivered too late to contribute meaningfully to their projects. Clearly, there are many current and potential users in the NASA community who require translation services, but only if the information can be delivered in a timely fashion. MT could prove to be a significant resource and asset to such R&D users. NASA is taking the first steps by installing SYSTRAN software on mainframe equipment, to begin testing an MT system on a global network.

Department of Energy (DOE). Through exchange agreements with foreign governments and international organizations, foreign research information is collected and made available to DOE through its Office of Scientific and Technical Information, and to the public through the Commerce Department's National Technical Information Service. Bibliographic citations and abstracts in English of this literature are available in publications and in on-line databases, such as the Energy Science and Technology Database (EDB). Of about 200,000 document citations DOE processes annually, about 50,000 are available only in foreign language text— predominantly Russian, German, and Japanese.

Initially, DOE had an active MT program. In the early 1960s, efforts focused mostly on Russian and Middle Eastern language translations. Today, no major MT activity exists in DOE.

Last year, there were over 200 requests for translations at each of the DOE National Laboratories, but fewer than half were actually performed. As with NASA, DOE scientists often cancel their requests for translations due to delays. Over 80 percent of the translation requests are for Russian, German, Japanese, and French. If quality machine translations were available in a timely manner at a reasonable cost, DOE scientists might request and use translations more often.

National Institutes of Health Library. Administered by the National Institutes of Health under the Department of Health and Human Services, this facility has two full-time staff who translate from Russian, German, French, Spanish, and Italian into English, and occasionally from English into French or Spanish. It also contracts additional translation services, mostly for Japanese, Chinese, Finnish, and Swedish. This unit has received increasing requests for Japanese and Chinese translations in recent years. If requestors lack funds for contract translations, they either wait for the in-house translators or drop the requests.

In the last four years, translation activity has been stable. But in-house translation has decreased, while dependence on contractors has increased. This is due not to a lack of demand but to a lack of needed clerical and administrative staff support. Yearly translation totals for contractors and staff are 800,000 to 900,000 words. When a published work is translated, a copy of the translation is sent to the National Translations Center of the Library of Congress, but no record is kept of translated articles. The use of MT could be cost-effective and promote rapid translation and dissemination of research information from abroad on such important and time-sensitive subjects as AIDS and other infectious diseases, cancer, environmental health, neurological disorders, and infant mortality.

National Translations Center (NTC). Under the Library of Congress, NTC provides no translation services. It serves as a clearinghouse for translations done throughout the United States. It receives deposits of full-text translations and titles of works translated, which it shares with its subscribers when they request searches. Each year, NTC receives over 12,000 translations from public-sector and private-sector organizations. Over 75 percent of these deposits are from American businesses, and most of the collection is S&T-oriented. NTC has a relatively small budget and had a manual indexing system. It expects to recoup most of its costs through user fees for its products and services. Recently NTC putting descriptions of current document acquisitions into machine-readable format and distributing tapes of these records without charge to information centers. Augmented use of MT among NTC's translation sources could amplify the effectiveness of the NTC collection.

Federal Research Division (FRD). Also under the Library of Congress, FRD employs 40 analysts who, on a reimbursable basis, provide "data briefs" to U.S. Government agencies on subjects of interest; perform translations as part of directed research (not as a separate service); and extract and abstract information from articles, conference proceedings, book chapters (rarely an entire book), and technical reports. They evaluate literature in nearly 25 languages, mostly Russian or Eastern European languages, followed by French, German, Portuguese, and Scandinavian languages. Generally, after clients receive translated citations and abstracts, they decide whether more needs to be translated. Thus, FRD provides two services: crude translations and factual information. Each month, it produces about 75 abstracts and 100 citations. With recent budget cuts, FRD staff lost a substantial Asian language capability, but retain some Chinese and Japanese translation skills. Before the cutbacks, FRD had over 200 employees producing over 10 times the current output. The resulting 40,000 to 50,000 citations are maintained in an on-line database. Access to MT text could help offset staff shortages in meeting clients' demands for services.

Office of Language Services (OLS), Department of State. Like FRD, OLS provides translation services to other Government agencies on a reimbursable basis. Requests for translations come from such agencies as the National Institute for Standards and Technology, National Oceanic and Atmospheric Administration, and Smithsonian Institution.

OLS has a unit for interpretation and another for translations. The latter has about 20 full time staff, but uses hundreds of contract translators. Within this unit are four sections, devoted to Romance-language translations into English; English into Spanish; French into English; and Russian, German, and all other languages into English.

Generally, geopolitical activity most affects which languages are in greatest demand at OLS. Events in Eastern Europe created an unprecedented demand for Russian and East European-language translations. In recent years, requests from other agencies have declined, due largely to reduced budgets for translations, not to a lack of demand. An MT capability could help lower the costs of human translation and enhance OLS' capacity to fulfill agencies' increasing needs.

U.S. Government Support for MT Development

History. While the seeds for MT development were generated in Europe, it was in the United States that the movement flourished in the 1950s and early 1960s. After the Sputnik spacecraft's launch in 1957, MT projects began at some 20 institutions nationwide. Among them was the University of Washington, which developed a Russian-English system that was later picked up by IBM and installed at the Air Force's Foreign Aerospace Science and Technology Center under the name of Mark I. This system used the simplest kind of "direct" approach, consisting of little more than a word-for-word lookup. It had so many problems that it was soon scrapped in favor of Mark II, which had somewhat more linguistic capability. However, the upgrade was not sufficient, and in 1969 Mark II was abandoned in favor of SYSTRAN.

Even in those early days, more sophisticated approaches were being explored at Georgetown University, MIT, and the University of Texas at Austin. Georgetown's research, funded by a CIA grant of $1 million, had a theoretical linguistic framework (drawing on Tesniere and Zellig Harris) but was also "empirical," in that development was based on incremental mastery of a large corpus of naturally occurring text. At MIT and Texas, more purely theoretical approaches were used, in which linguistic theory determined the system's course of development.

By the early 1960s, expectations were high, and the various projects promised in all sincerity that "fully automatic high-quality machine translation" (FAHQMT) was just around the corner. The funding agencies, however, began to get impatient and ask hard questions about what was happening with their money. These questions led the National Academy of Sciences in 1964 to appoint a team of six linguists, the Automatic Language Processing Advisory Committee (ALPAC), to look into the matter. Their report, which appeared two years later, concluded that since FAHQMT was impossible, the technology could never replace human translators. They said that funds would therefore be better spent on basic linguistic research and machine aids for translators.

While most MT activity declined sharply in the United States after 1966, Georgetown's Russian-English system (GAT) and Texas' German-English system (METAL) survived. Not all Government funding of MT projects stopped for the next 25 years, as is commonly believed. In fact, without the Government's support of MT research during this period, the three MT systems that dominate the Western Hemisphere market— SYSTRAN, LOGOS,and METAL— would probably not exist today. The GAT system was used by the Atomic Energy Commission to scan technical material at Oak Ridge National Laboratory for over a decade. METAL, revamped later under the sponsorship of Siemens AG, is now marketed commercially, and used with success in Europe. LOGOS also received some early Government support.

SYSTRAN. Beginning in 1968, Dr. Peter Toma, a Hungarian-American who had gained practical MT experience from working on the Georgetown project, received support from the U.S. Air Force for his SYSTRAN system. In 1969, the system was installed at the Foreign Technology Division (FTD is now the Foreign Aerospace Science and Technology Center, FASTC) at Wright-Patterson Air Force Base, where it has translated Russian S&T documents since 1970 and German and French documents since 1987. By 1983, the Russian-English system had attained 85 percent accuracy. FTD then began gradually reducing its support of Russian-English in favor of German-English and French-English language pairs (which were initially funded by non-U.S. Government sources). Thus, of the $6 million in these funds for SYSTRAN over the last 20 years (at a fairly steady rate of $300,000 a year), $4 million was invested in Russian and approximately $2 million in German and French. These three language pairs were made directly available to researchers and analysts (usually non-translators) at FTD's 1,400 terminals in 1987. Since then, FTD (now FASTC) has installed SYSTRAN at over 10 additional Government sites and has downloaded the interactive system to run on various configurations of stand-alone IBM PCs, so that it can be used at remote sites lacking access to an IBM mainframe.

Steady U.S. Government support for the SYSTRAN Russian-English system has created a solid framework for today's SYSTRAN, which translates among all major Western languages, as well as into and from Japanese. Experts view as invaluable the basic system software, research and diagnostic tools, immense dictionaries and indexed databases of actual text (6 million words of Russian alone), and years of experience in creating MT systems for real-world users— which all evolved from Government support. Also, funding from NASA to develop an English-Russian system during 1973 to 1974, to translate documentation for the Apollo-Soyuz project, led directly to today's eight operational language pairs, with English as a source language, and to the system that enables the Xerox Corporation to market its products abroad six months earlier than before.

In 1980, SYSTRAN's experience and linguistic resources were used to create the more difficult Japanese-English system, which received initial funding of $2.5 million between 1980 and 1984 from a Japanese corporation. When this corporation granted the U.S. Government the right to develop and use this language pair in 1985, Latsec (the U.S. SYSTRAN developer) received an additional $300,000 a year (mainly from FASTC) for Japanese-English dictionary development and linguistic programming. The Japanese-English system, which had already proved adequate for translating technical manuals in a study done for Xerox in 1983, is installed for testing at the Foreign Broadcast Information Service, where it will initially be used to translate titles for information scanning. The system's parser is already adequate for this task, but more funding is needed to upgrade the system's dictionaries (currently less than one-sixth the size of the Russian dictionaries) in pertinent technical areas. During the next two years, progress in dictionary-building for this system should be accelerated if it is to meet the U.S. Government's needs.

The Spanish-English system is the focus of a small-scale development effort for DOD. Other projects awaiting sponsors include a pilot Chinese-English system demonstrated to Government representatives in the early 1970s, and Korean-English pairs developed as prototypes in 1988.

LOGOS. Another U.S. company which got its start through post-ALPAC Government funding is Logos. In 1969, the Air Force gave Bernard Scott, a former Air Force intelligence specialist, a chance to test his theories on MT by developing an English-Vietnamese system to translate thousands of military training manuals. Although the system proved its value by translating several million words of technical English during its year of existence as a production system, the end of the Vietnam conflict terminated Government funding. Still, the experience left Logos with a useful English analysis module, which was expanded into a working prototype of an English-Russian system for the CIA in 1973, but funding never materialized. Next came the unfortunate experience of developing an English-Farsi system for the Iranian government just before the Shah's fall from power. After three discouraging experiences with less marketable language pairs, Logos decided to work on Western languages, and by 1988 had penetrated the European market with its German and French systems. By 1989, the LOGOS system was installed at over 40 sites in Europe and North America, and had been selected by the Canadian government for a plan to introduce MT into all government agencies. Yet, without more funding sources to develop new language pairs, the Logos staff has been reduced to a skeleton crew, and the company has withdrawn support from its European customers in order to concentrate on the more promising Canadian market.

METAL. The third example of a U.S. Government-funded MT system has followed a very different pattern. Unlike other operational MT systems, METAL is the outgrowth of 20 years of solid research in theoretical linguistics at the University of Texas from 1959 to 1979. Between 1962 and 1974, the Linguistics Research Center (LRC) in Austin received about $325,000 a year from the Army, Air Force, National Science Foundation, and other agencies to do research on the German-English language pair. In 1978, the LRC received a grant from the Rome Air Development Center to study the feasibility of an operational German-English MT system. About this time, the German firm Siemens, which was seeking a means of translating its own technical manuals, began providing support to the LRC; by 1980, Siemens had become the sole support for the development of an operational German-English system. By 1988, the system was ready for marketing in Europe, where it is installed at over 20 sites. Marketing in the United States is expected soon. Much of the development work has moved to Europe— a case where the profits from technology developed with U.S. Government funding will go to a non-U.S. company.

Pan American Health Organization (PAHO). An example of a successful non-commercial system which has received some U.S. Government funding is ENGSPAN, the English Spanish system developed at PAHO. Originally designed for internal use, ENGSPAN was partially supported by a grant from the U.S. Agency for International Development (AID), which enabled it to incorporate enhancements based on contemporary linguistic theory. ENGSPAN became operational in 1985. Besides PAHO and AID, other installation sites are in Colombia and the Philippines, to fulfill its sponsor's goals of disseminating information on health and agriculture to Third World countries.

Intelligence Community Staff (ICS). ICS has supported a project to develop an optical character recognition system for Japanese, with total funding of about $175,000 to one U.S. firm. Managed by the Foreign Broadcast Information Service, this project just saw comple-

tion of a prototype with a recognition accuracy of 85 percent. ICS hopes to continue this effort under wider Government auspices.

National Science Foundation (NSF). NSF supports basic research on MT, including fundamental work in computational linguistics and natural language understanding that form the core technology for MT. In the last few years, NSF has also supported specific MT projects at a level of approximately $350,000 yearly. In FY 1990, the Computer and Information Science and Engineering Directorate (CISE) funded three projects: "Machine Tractable Dictionary as Tools and Resources for Natural Language Processing," at New Mexico State University; "Multilingual Natural Language Processing," at Carnegie Mellon University; and "A Sub-Language Approach to Japanese-English Machine Translation," at New York University. In addition to the fundamental work and specific MT projects, CISE supports research in a number of related technologies: optical character recognition, text processing, and speech recognition and understanding aimed at more efficient input and output for MT systems. At the international level, NSF co-sponsored the Japan Technology Evaluation Center (JTEC) team's 1990 visit to Japan to survey MT research and is exploring joint research with the European Commission.

Department of Defense (DOD). Under the Defense Advanced Research Project Agency (DARPA), the Software and Intelligent Systems Technology Office (SISTO) has recently initiated a multi-million dollar project to develop MT systems for several source languages into English. SISTO has supported the Japanese Translation Project at the Courant Institute of New York University ($1.1 million), which investigates commercially available Japanese-English and English-Japanese MT systems. Due to budget constraints, however, the project is on hold and might be dropped.

At the Defense Technical Information Center (DTIC), the Defense Gateway Information System (DGIS) is providing access to numerous foreign databases (and, soon, to on-line MT facilities). With this capability, the database user will be able to make informed decisions about subscribing to foreign databases and get a partial translation to determine if a document needs full translation by human translators. In the same way, it should be possible to evaluate and translate search terms into another language in real time. The knowledge base to supply these equivalent terms could be built upon the *DTIC Thesaurus*,which is being adopted as a NATO standard. DGIS will include the SYSTRAN MT system as a prototype for an on-line translation service.

Several MT activities have been completed or will be initiated by other DOD agencies. These include a Korean-English system at Rome Air Force Base, TACCINS at Ft. Monmouth Army Base, and JNIDS for the intelligence groups. Details of these projects are classified.

National Security Agency (NSA). For several years, NSA has supported the Center for Machine Translation at Carnegie Mellon University with annual funding levels of $60,000 to $80,000. It has sent four researchers there to study MT technology and is considering future MT projects.

National Aeronautics and Space Administration (NASA). NASA's Scientific and Technical Information (STI) Program views MT as an integral element of networked and distributed S&T information-processing and information-access systems that will be established during the 1990s. These global networks will aim to improve the effectiveness and efficiency of NASA's R&D programs. Information technology has significantly affected the way research is conducted, and will continue to play an important role in that process. MT is one aspect of this technology that will have a major impact on multinational relations and information exchanges through real-time electronic mail translation, on-demand document translation, and multilingual bibliographic announcement services (such as NASA Scientific and Technical Aerospace Reports [STAR]). In an effort to provide better and faster access to translation services in these areas, the STI Program installed SYSTRAN MT software on mainframe hardware and has begun testing an MT system on a global network.

MT Development and Implementation Abroad

In Japan

Research in Japan on MT began around 1956, but dynamic growth in Japan's trade with the West caused burgeoning growth during the 1980's. The difficulty of the Japanese language for foreigners, and the need for a more efficient way to process the volume of S&T information available in English, led the Japanese to view MT as a technology key to Japan's growth as a dominant economic power. By 1987, 11 major computer or electronics firms in Japan had started an MT development project. The 1989 report of the Japan Electronic Industry Development Association (JEIDA) listed 25 Japanese organizations as developers of MT systems, 14 of which were said to be marketable or at least operational.

The Japanese government has played a significant role in this development by supporting the national MT project (Mu) at Kyoto University from 1982 to 1986. This project, which included cooperative research with manufacturing firms, demonstrated the feasibility of developing a large-scale MT system in Japan. It was taken over in 1986 by the Japan Information Center of Science and Technology (JICST), which now uses Mu 2 to translate both Japanese and English abstracts. Since 1986, Japan's Ministry of International Trade and Industry (MITI) has funded the Center of the International Cooperation for Computerization (CICC) consortium to develop translation between Japanese and the languages of neighboring countries. It also provided seed money for a project of the Electronic Dictionary Research Institute, Ltd., to create an electronic dictionary for English and Japanese, and for the ATR project, to develop a system for automatic interpretation of telephone calls between Japanese and English. Both projects are supported by industry consortia.

According to the JTEC study team, which visited 15 Japanese MT research sites, Japan leads the United States in funding R&D in MT, in commercial use and general acceptance of MT, and in integration of MT into the office environment. The United States is strong in basic research on natural language processing.

One of the few areas in which the United States leads is the diversity of language pairs covered by R&D in MT. Currently, 90 percent of Japanese MT efforts are focused on translation between English and Japanese, with 10 percent focused on translation between Asian languages. Relatively little effort is devoted to translation into other Western languages. The U.S. SYSTRAN system has more language pairs (roughly 25) in use or under development than all Japanese firms combined. However, the Japanese can be expected to add more language pairs as their products penetrate other commercially attractive markets. Fujitsu is expanding into European language combinations with its work on Japanese into French, German, and Spanish. It will extend its project to translate technical and economic documents for the European Commission (EC) by offering post-edited Japanese-English translations to other European users for the same price that the EC has been getting (about 12 cents per word). Sharp is opening an MT laboratory in England.

In number of customers, the market leader is believed to be Fujitsu, which has marketed its Japanese systems since 1985 and reportedly sold over 100 systems by 1989. It is not clear how many of these systems are in use; the best-known application is at the Mazda Motor Corporation.

Although most Japanese MT development originated in Japan (with underpinnings of U.S. basic research on theoretical linguistics), the Systran Corporation of Tokyo developed its system from U.S. technology. Building on U.S. efforts which it had financed from 1980 to 1984, it continued to develop its Japanese-language pairs in Japan from 1984 to 1990, bringing the English-Japanese pair to a fully operational level. The incorporation of a 250,000-term S&T dictionary and a 250,000-term medical dictionary into the English-Japanese dictionary made this the world's largest operational MT dictionary, with 414,000 terms and expressions. As mentioned earlier, 1985 saw the agreement by Systran's Japanese owner to allow Latsec (the U.S. SYSTRAN developer) to continue to develop the Japanese-English system under its existing contract with the Air Force, and to grant the Government a license to use the resulting system to perform translations for any U.S. agency.

The JTEC team found that a major application of MT systems in Japan is to translate technical manuals for Japanese electronic products sold overseas. IBM and Sharp use their systems internally to produce manuals. Oki has linked its PENSEE translation system to its electronic-mail system for use within the company. Fujitsu's ATLAS II (Japanese-English) is available on the information utility NiftyServe, where it gets 50 to 60 interactions per day at a cost of about 7 cents per minute plus less than 1 cent per word. NHK uses Catena's STAR system (English-Japanese) to monitor AP wire reports, while CSK's ARGO system specializes in translating financial reports and stock-market information. Two large translation bureaus, Inter Group and IBS, use MT systems: Inter Group uses Fujitsu's ATLAS II to translate abstracts for the European Commission.

For many years, Japanese developers have insisted that machine translation of Japanese cannot succeed without pre-editing of the input text--usually a complex process requiring knowledge of both Japanese and the MT system. This is significant; it means that U.S. use of most of these systems, even for information scanning, could require people who know Japanese well. But it is precisely because of our lack of Japanese-language capability that many Americans are interested in these systems. Until now, SYSTRAN has been the only Japanese system that does not use pre-editing. Fujitsu is test-marketing its MT system as part of an information-retrieval system that allows the user to input an English keyword. This word is then automatically translated into Japanese and used to search the Japanese database; the resulting abstracts are translated by machine, using an automated pre-editing phase, and returned to the user on-line.

In the end, Japanese developers know that the performances of their respective systems do not differ greatly. Therefore, they aim to enhance their products' competitiveness by providing user-friendly tools for pre- and post-editing MT texts and creating and updating customized dictionaries. Many MT developers are devising optical character recognition devices for other purposes as well. Until now, these user interfaces have not been designed for monolin-

gual English-speaking users. Also, the MT systems are usually tied to Japanese hardware and operating systems (and thus often serve as a door-opener for the developer's own hardware). These barriers will have to be overcome before the systems can penetrate U.S. and European markets. In 1989, Geoffrey Kingscott stated in a report to the Commission of the European Communities (CEC) that "the Japanese are convinced that the way to future success in all domains lies through machine translation." If Kingscott is correct, these barriers will not exist for long.

This point was reinforced by Makoto Nagao of Kyoto University, often called "the father of Japanese MT," in response to a question raised at the symposium on Japanese-English MT held at the National Academy of Sciences in 1989. Alluding to the U.S. Government's role in promoting the development of MT, Nagao wrote that a "view for the promotion of machine translation and its related technologies is that the natural language processing technology is a key technology in the future information society. Therefore, the Government must promote R&D in this area. Otherwise, the country will be defeated by others in the information war."

In Europe

The MT market is dominated by two large general-purpose systems, SYSTRAN and METAL, both developed in the United States and modified in Europe. In 1976, the CEC acquired the rights to use and develop SYSTRAN for all of its 72 language pairs and to market it to public agencies of its member states. In addition, SYSTRAN development centers in Paris and Luxembourg, privately owned by the Gachot Group, market the system to industry and all non-member states, and perform dictionary and linguistic development in coordination with the U.S. SYSTRAN companies, Latsec and Systran Translation Systems.

The CEC's annual investment in SYSTRAN has grown from $75,000 in 1975 to $2.5 million. A team of 35 people support the development of 16 language pairs; 10 of these pairs are fully operational, with one or two new pairs entering development each year. Initially, CEC development work was limited to dictionaries, but it has grown to include all aspects of the system. New language pairs, however, are still developed at Systran Translation Systems in California. In the past decade, information has been informally exchanged between the U.S. Air Force and the CEC, as major system developers, under Latsec's control. A more formal agreement between Latsec and the CEC led to Latsec's acquisition of the CEC dictionaries in 1989. The CEC is about to purchase a six-month license to test (but not develop) Systran's Russian-English system for use within the CEC.

Six of the 30 directorates-general at the CEC use SYSTRAN daily for translating reports and minutes of meetings. SYSTRAN's extension to other directorates is limited by the fact that only 1,000 of the 7,000 translators in Brussels have access to equipment that is compatible with SYSTRAN's infrastructure. Outside the CEC, the biggest public-sector users are NATO (30,000 pages a year), the Nuclear Research Center in Karlsruhe, and Aerospatiale.

The CEC reports that SYSTRAN has made possible a 50 to 60 percent cost savings in translating documents that require post-editing, and an 80 percent savings for users of raw translation. It considers SYSTRAN to be a "universal system" that can handle 95 percent of the CEC' s subject fields and text types with its highly developed, multi-target dictionaries; outside users can add "local dictionaries" for their specialized terminology.

Now winding down after six years, the $30 million Eurotra project was the biggest MT undertaking in history, both in number of people involved and probably in funding level. It aimed to provide rapid translation among all member nations' languages, through the interlingual approach. For political reasons, its development was divided among all EC nations. There was a small central coordinating unit, but the administrative complexity probably diminished the project's effectiveness. While no operational MT system resulted, even in limited prototype form, a number of efforts in computational linguistics research were fostered. Kingscott expressed the opinion that "the combination of the stimulus provided by EUROTRA research, and the practical experience gained from SYSTRAN, gives the Commission of the European Communities a central role in machine translation, and with it a heavy responsibility."

LOGOS— a general-purpose American MT system almost as old as SYSTRAN achieved major marketing success in Europe, beginning in 1985, with its German and English source systems. Its clients included Nixdorf, IBM, and Hewlett-Packard. By 1988, it was reported to control 35 percent of the European MT market. Recently, however, METAL has been gaining an increasing share of the European market.

Several important research systems developed at European universities have not been as successful as METAL in achieving commercial status. GETA (Groupe d'Etudes pour la Traduction Automatique), the MT research group at the University of Grenoble, has conducted government-sponsored research since 1961. It produced what MT historian John Hutchins described as "the most advanced of current MT systems," but its attempts at commercialization have foundered. The group is now working on a project (LIDIA) to assist writers in preparing texts that are more suitable for MT. The NTRAN project at the University of Manchester aims at an English-Japanese system for monolingual English-speakers; grammatical and semantic ambiguities in the source text are solved through dialogue with the user. The University of Sheffield is working on a Japanese-English system. Both Manchester and Sheffield are exchanging researchers with Japan.

TITUS 4— the fourth version of a system designed to translate textile abstracts written in a controlled language, is used only by the French textile industry. There seems to be no deadline for commercializing the ROSETTA system, an experimental project of Phillips, a Dutch firm in Eindhoven. It is one of the few systems-research efforts by private industry outside of Japan. Work is concentrated on Dutch-English and Dutch-Spanish language pairs. DLT (the Distributed Language System), an interesting project administered by BSO of Utrecht, which receives support from the CEC and the Dutch government, needs several more years of work to overcome the difficulties of using Esperanto as a pivot language for translating among multiple languages. WINGER is a Danish system, developed by Winger Holdings

A/S as a result of innovative data-storage techniques invented by the company. The system requires interactive text analysis by an operator during translation, but is in use for English-to-Dutch translation.

Until now, systems produced in the United States and modified in development centers there and in Europe have dominated the European market. Geoffrey Kingscott has emphasized the CEC's central role in MT and warned: "If Japan is not to dominate the next three decades of machine translation activity, as the United States has dominated the last three decades, major coordinated efforts will have to be made. Now that research and development in the United States seems to be proceeding on a scattered basis, it is in Europe that major non-Japanese initiatives may be expected."

In Other Countries

Canada. The best-known example of a translation system designed for a specific purpose—and one of the most successful MT applications anywhere— is METEO. This system has been translating Canadian weather forecasts from English into French since 1977 with little human intervention. In 14 years of operation, it has translated 100 million words, currently 10 million per year; less than 3 percent of its output requires post-editing. Developed by the University of Montreal, the system was the result of long-term funding by the Canadian government. In 1988, a French-English version was introduced. Not only has METEO been cost-effective, but it also has relieved the monotony of translation work and thus eliminated high turnover in the staff.

METEO has its roots in research at the University of Montreal. Another of the University's projects, TAUM-AVIATION, envisioned the translation of aircraft maintenance manuals. Begun in 1977, it was eventually abandoned.

Although the Canadian government has a pressing need for translation, it did not develop any practical MT applications again until the late 1980s. In the meantime, it developed an enormous terminology data bank called TERMIUM, which contains 900,000 terms in 50 subject fields in English and French.

In 1989, the government chose LOGOS as the general-purpose MT component in its plans for a network-based, government-wide automation of translation. However, due to lack of steady funding, Logos has been unable to maintain sufficient staff for customer support. During the last two years, Canada's Office of the Secretary of State has evaluated several MT systems for purchase. Also, a translation company called Lexi-Tech was formed to use MT to translate 100,000 pages of technical manuals on shipbuilding for the Canadian Navy.

XLT is a system developed in Canada by Socatra, a translation service bureau, for translation from English into French. A PC version of the product was launched on the market in late 1990. Throughput of the system is reported to be 60,000 words per hour.

USSR. Soviet research on MT, begun in 1955, underwent a period of disillusionment in the mid-1960s. Progress was also slowed by limited access to computers. Since 1974, activity has been concentrated in the Center for Translation of Scientific and Technical Literature and Documentation in Moscow, where general-purpose systems have been developed for translating English, French, and German into Russian. With the dissolution of the USSR, the future of this activity is unclear.

China. There is little information on early MT research in the People's Republic of China. At MT Summit I, it was reported that 15 groups in China were doing research on English-Chinese systems. China is one of the countries engaged in joint MT research with Japan's Center of the International Cooperation for Computerization (CICC) consortium. A system called CULT, to translate between English and Chinese, was developed by the Chinese University of Hong Kong and implemented in 1969. The system is noted for its pioneering work in interactive pre-editing. JFT-IV is the current version of a system which has been under development in China since 1976. It is intended for translation into Chinese from English, French, German, and Russian, with most of the work so far concentrating on English. An experimental model is being developed to translate from Esperanto to Chinese.

Korea. Several Korean universities and research organizations have been developing MT systems in cooperation with foreign counterparts. The Systems Engineering Research Institute (SERI) of the Korean Advanced Institute of Science and Technology (KAIST) began a government-financed development program in 1983, working with Fujitsu to develop Japanese-Korean systems, and with France's GETA to develop English-French and English-Korean systems. Korean-Japanese is also under development. Other projects have been undertaken in cooperation with IBM, Japan's Waseda University, and NEC of Japan.

Malaysia, Indonesia, and Thailand. These countries are working on projects to translate between their languages and Japanese as part of the Japanese government's CICC project. The Malaysian Institute of Science, along with France's Grenoble University, is pursuing a project to translate English into Malaysian.

Israel. The TOVNA system was launched at a conference in London in 1987. However, its Israeli developer is said to have worked on it throughout the 1970s. TOVNA claims to be the only MT system that learns from its users, i.e., it remembers users' changes and incorporates them into future translations. The first language pair, English-French, has been installed as a pilot system at the World Bank; English-Russian and French-English are under development.

Bolivia. ATAMIRI is an MT system developed by Ivan Guzman de Rojas, a Bolivian mathematician. It uses the Indian language Aymaraas, an intermediate syntactic representation for translation into multiple languages. Wang International Translation Centers employ ATAMIRI to produce technical manuals.

Roles the U.S. Government Could Usefully Play

At the MT Summit held in Washington, D.C. July 1-4,1991, the Association for Machine Translation in the Americas (AMTA) held its first meeting. During the conference, the MT Working Group visited a meeting of the AMTA. Members of the AMTA board offered some views about the role of Government and cooperative efforts with the MT industry. These views are reflected in this report, and have led to identification of the following options.

Increase support for research and development of MT.

Sponsor Research. Major improvements in MT technology are possible, but they will require sustained research and innovative ideas. The Government could speed the process and underwrite some of the risks by increasing its funding levels for long-term MT research in universities and industry. The Government might also provide some SBIR (Small Business Innovation Research) funding.

Sponsor Enhancements. In order to meet validated Government needs, the Government could pay for the enhancement of one or more currently available MT systems. While this effort might not have as much long-term impact as real research, it could provide some nearer-term benefits and broaden Government use and access to the underlying technology.

Evaluate Performance. Methods for objectively evaluating the performance of MT systems would provide useful information to potential buyers of existing systems and would help researchers develop better systems. The Government could help industry develop standard evaluation methods for commercial MT systems. The Government could also sponsor periodic evaluations of commercial and research systems.

Forge stronger linkages for transferring knowledge from basic research to products.

Mechanisms might include SBIR, CRADA (Cooperative Research and Development Agreement), ATP (Advanced Technology Program), and patent licensing. MT developers would be appropriate candidates for all of these programs.

The ATP Model. ATP was created through the Technology Competitiveness Act of 1988 to support U.S. companies in developing precompetitive generic technologies with significant commercial promise. Administered by the National Institute of Standards and Technology (NIST), ATP emphasizes enabling technologies with strong and broad potential. Grants are awarded to companies, alone or in partnership with universities and research institutes. In 1991, NIST announced a second round for proposals, including those for developing computer software. MT might well qualify for such support, especially in creating evaluation standards, multilingual dictionaries, and large corpora of text in a variety of languages and subjects .

Sponsor Workshops. To facilitate technology transfer between researchers and system developers from different organizations, the Government could sponsor a series of technical workshops. These meetings would foster information exchange and possibly collaborative arrangements.

Target U.S. MT industries for invitation to Federally sponsored trade shows. To demonstrate to U.S. businesses the role that language and cultural awareness can play in boosting their international competitiveness, U.S.-sponsored trade shows could ensure that MT companies and trade associations are invited.

Devise mechanisms for pooling and leveraging the necessary resources among agencies — personnel, programs, and monies.

Cooperative Procurement. One approach is a "cooperative buy," featuring a centralized procurement of translation services for several agencies. Benefits would include not only economies of scale, but also the use of a machine translation service as a focal point for filtering the massive flow of foreign data and quickly "brokering" the right information to the right agency experts.

Assess Needs. A study that carefully analyzed public-sector and private-sector needs for MT would help system developers understand where to invest their energies. The study could identify the language pairs, domains, and interaction methods of interest today, project those into the future, and quantify the demand.

Establish a Clearinghouse for MT Information. Potential buyers and sellers of MT systems would be aided by the existence of up-to-date lists of MT systems and system developers. Accurate information about the strengths and limitations of MT technology in general and, where available, performance characteristics of specific systems would also be helpful.

Provide Linguistic Data. The development and evaluation of MT systems would be facilitated by the ready availability of certain kinds of linguistic data, especially large quantities of parallel texts in different languages in electronic (character-encoded) form. The Government could obtain existing parallel texts and distribute them in a standard format. The Government could also obtain or produce lexicons and grammars that have broad utility.

Conclusion

The world is now well into the Information Age and stands at the threshold of major advances in natural language processing. MT can be an important part of those advances, and progress in MT will help fuel progress in other areas of natural language processing.

In the early 1960s, the United States led the world in MT research and development, and several of the best commercial systems today evolved from U.S. work. In the past decade, however, Europe and Japan have invested much more heavily in MT, and they could easily come to dominate the world market. This trend would have three negative consequences.

First, the market for MT, though not huge now, will grow substantially as international trade increases and as the quality of MT systems improves. The level of indigenous MT capability in the U.S. could have a direct bearing on its ability to compete internationally.

Second, the market for other kinds of natural language processing will be enormous— so much so that MT should be considered a critical information technology. Those who are investing heavily in MT will be able to leverage their results and insights to advance their monolingual natural language work.

Third, the Government itself has significant and increasing needs for translation, and for economic reasons will have to depend more heavily on MT. Yet Government needs do not necessarily match major commercial needs (in terms of languages or domains), and the Government may not be able to get away with just buying standard products. In some cases, it will have to tailor those products (or the underlying technology) to its own special needs.

The good news is that the United States has considerable technical strengths and could be a very strong contender both in MT and in all areas of natural language processing. However, some degree of Government investment and encouragement will be needed to offset the huge investments made by foreign governments and corporations. Fortunately, that investment need not be so large to be effective, and it is not too late to have a major impact.

The United States is entering a challenging and dynamic era. Conquering the language barriers to information generated abroad could help strengthen the Nation's presence in worldwide developments. In that endeavor, machine translation technology could have a role to play for America's economic vitality and for the quality of life of its citizens. Potentially, the Government's investment in this effort could catalyze the Nation's progress in the Information Age.

APPENDIX A: Membership of Working Group on Machine Translation FCCSET Committee on Industry and Technology

Department of Commerce
Joseph E. Clark, Chairman
Phyllis Genther-Yoshida, Co-Chairman
Boyd Alexander
Tim Feinstein
Victoria Kader
Mike Keplinger
Tom Kusuda
David Shonyo
Jack Williams

Department of Defense
Robert Billingsley
Charles Wayne

Department of Energy
Elizabeth Buffum
Wanda Ferrell
Norman Kreisman
Dora Moneyhun

Department of Interior
Herman Enser

Department of State
Linda Staheli

Foreign Broadcast Information Service
Wayne Kiyosaki

National Science Foundation
Y.-T. Chien
Emily B. Rudin

Office of Science and Technology Policy
Jennifer Bond
William Whyman

We gratefully acknowledge the comments of:
Joann P. Ryan
Muriel Vasconcellos
Jennifer DeCamp
Sergei Nirenburg
Richard Lambert
Martha Harris

APPENDIX B: Selected Sources On Machine Translation

Recent Bibliographies

[AGARD] Advisory Group for Aerospace Research and Development. 1990. "Bibliography." In: *Benefits of Computer Assisted Translation to Information Managers and End-Users.* Neuilly-sur-Seine: AGARD, North Atlantic Treaty Organization. AGARD Lecture Series No. 171. pp. B1-B34. Useful summaries are provided in English or French for some 270 recent citations on MT or closely related technology, most of them papers from conferences (e.g., 22 cites from Coling '86, q.v.).

King, Margaret. 1987. "Bibliography." In her: *Machine Translation Today: State of the Art.* Edinburgh: Edinburgh University Press. pp. 391-435. A total of 665 unannotated but selected entries are divided into five sections: MT up to 1973(57 entries), MT 1973 onward (461), software (45), linguistics and computational linguistics (66), and artificial intelligence (36).

[NTIS] National Technical Information Service. *Machine Translation: Foreign Language Translation and Natural Language Understanding*, citations from NTIS Database, Jan 1970-Jul 1989. Springfield, VA. A total of 126 full database citations are given on a variety of MT- related topics.

Slocum, Jonathan. 1987. "A Machine(-aided) Translation Bibliography." In his: *Machine Translation Systems.* Cambridge, etc.: Cambridge University Press. pp. 265-341. Some 850 "currently accessible documents" in English, French, or German from 1973-1986 are cited but not annotated. While there is much overlap with the bibliography in King (1987, q.v.), this list is more complete.

Journals

Applied Computer Translation. Annual subscription $60 individual, $120 library/corporate; apply to Sigma Press, 1 South Oak Lane, Wilmslow, Cheshire SK9 6AR, United Kingdom.

ACT began publication in 1991 under the editorship of Tony McEnery, University of Lancaster, and intends to be a quarterly. It encourages an interdisciplinary perspective by bringing together concepts from linguistics, computer science, and related fields in an easily understandable form. Topics covered include knowledge-based and probabilistic MT, with emphasis on applications.

Computational Linguistics. Individual subscription available through membership in Association for Computational Linguistics ($25 per year); apply to Donald E. Walker, Bell Communications Research, 445 South Street, MRE 2A379, Morristown, New Jersey 07960. Institutional subscriptions $60 a year; apply to MIT Press Journals, 55 Hayward Street, Cambridge, MA 02142-1399.

LL, the ACL's official quarterly journal, publishes theoretical papers on natural language processing in general, and frequently reviews milestone publications in the MT field. It is a refereed journal. Subscription includes a newsletter supplement, *Finite String*, which contains site reports and an up-to-date calendar of events.

Electric Word [formerly *Language Technology*]. P.O. Box 5477,1007 AL Amsterdam, The Netherlands. Claimed by editor Geoff Pogson to be "the world's least boring computer magazine," E~ made a valiant attempt to appear bimonthly and keep up-to-date on software and personalities in MT, among other topics, but has not published an issue since August 1990.

Language Industry Monitor. Annual subscription $95. Apply to LI Monitor, Eerste Helmersstraat 183, 1054-DT Amsterdam, The Netherlands. *LI Monitor* appeared for the first time in February 1991 and will be published bimonthly under the editorship of Colin Brace. Its appearance is slim (8 pages) and solemn (no advertising), but looks should not deceive; a lot of substance is packed into the brief articles, which are also highly readable. Rather than offering in-depth illustrated stories, the *LI Monitor* will concentrate on facts and news of products and events in language processing, with emphasis on MT. It will no doubt get fatter as it becomes more successful.

Language International. Annual subscription $57 individual, $90 corporate/library; apply to John Benjamins North America, 821 Bethlehem Pike, Philadelphia, PA 19118, or John Benjamins Publishing Co., P.O. Box 52519,1007-HA Amsterdam, The Netherlands. This bimonthly magazine was inaugurated in 1989 under the editorship of Geoffrey Kingscott. Every issue has a section on MT with up-to-date news and articles. In-depth interviews provide valuable background. As with its predecessor, *Language Monthly,* the coverage is largely European.

Newsletter of the British Computer Society. Natural Language Translation Specialist Group. Apply to Mr. W. Goshawke, 68 Barrington Road, Bexley-heath, Kent, England DA7 4DW. This publication appears sporadically and has carried interesting articles on MT over the last 10 years.

Machine Translation [formerly *Computers and Translation].* Annual subscription $48.50 individual, $109.50 plus $13 postage institutional; apply to Kluwer Academic Publishers Group, P.O. Box 358, Accord Station, Hingham, MA 02018-0358, or P.O. Box 322, 3300-AH Dordrecht, The Netherlands. *CaT*, a refereed quarterly, began in 1986 at the University of Texas/Austin under the editorship of W.P. Lehmann and the assistant editorship of Veronica Lawson. Initially, it was intended for MT users as well as re searchers, and contained some practical articles. In 1988, it moved to Carnegie Mellon University under the editorship of Sergei Nirenburg. Later, the name was changed. The material is now mainly theoretical.

MT News International. Newsletter of the International Association for Machine Translation. Available from the Association for Machine Translation in the Americas, 655 Fifteenth Street N.W., Suite 310, Washington, D.C. 20005. Contains current news of the association, conference reports, project reports, user views, recent publications, forthcoming events and related topics. A valuable current awareness tool.

Studies and Monographs

[ALPAC] Automatic Language Processing Advisory Committee. 1966. *Language and Machines: Computers in Translation and Linguistics*: A Report by the Automatic Language Processing Advisory Committee (ALPAC). Washington, D.C.: National Academy of Sciences, Division of Behavioral Sciences. National Research Council Publication 1416. In this much-cited study (funded by DOD/CIA/NSF), a panel of six scientific linguists concluded that "fully automatic high-quality" MT was impossible and recommended that investments be channeled instead into basic linguistic research and development of machine aids for translators. The premises are now considered out-of-date (see JEIDA 1989).

[FBIS] Foreign Broadcast Information Service. JPRS Report: Science and Technology Japan. July 25,1991. *Evolution. Applications of Machine Translation Systems.* Washington, D.C.: Joint Publications Research Service.

Goshawke, Walter, Ian D.K. Kelly, and J. David Wigg. 1987. *Computer Translation of Natural Language.* Wilmslow (U.K.): Sigma Press; New York, etc.: Halted Press. 275 p. The first of three essays in this volume, by Kelly, is a 70-page introduction to the current state of MT, mainly from the perspective of linguistic problems. The second, by Goshawke, is a description of the SLUNT number language, proposed as an intermediate language for MT. The third, by Wigg, illustrates the implementation of SLUNT in an International Communicator System (ICS) for translation between English and French.

Henisz-Dostert, Bozena, R. Ross Macdonald, and Michael Zarechnak. 1979. *Machine Translation.* The Hague: Mouton. Trends in Linguistics, Studies and Monographs 11. 265 p. The first of these three essays is Zarechnak's history, which contains valuable material based on direct personal experience. Macdonald explains the difference between theoretical and empirical approaches - a schism that persists to the present day and predicts an ultimate blending of the two. Henisz-Dostert analyzes the results of her survey of 58 MT users at Oak Ridge and EURATOM.

Hutchins, W.I. 1986. *Machine Translation: Past. Present. Future.* New York, London: John Wiley & Sons. 382p. This standard work on MT, soon to be updated, is noted for its thoroughness and impartiality. Hutchins, by profession a documentalist, brings scholarship to the task and emerges as the official chronicler of MT. Sometimes his reliance on published sources, rather than direct contact with MT developers, has resulted in some omissions or misinterpretation, but the new book is expected to correct any such problems in the 1986 edition. Hutchins is responsible for popularizing the notion of three "generations" of MT.

[JEIDA] Japan Electronic Industry Development Association. 1989. *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC. U.S.A.* Tokyo, JEIDA. 197 p. Under the chairmanship of Makoto Nagao, the

JEIDA Committee on MT, representing 20 MT developers, researched the state of the art around the world over a two-year period. The conclusions of this thorough study focus on changed circumstances since the 1966 ALPAC report. The 15 appendixes provide valuable raw data, including complete summaries of all MT systems in Japan and else where. If it has a weakness, it is that the estimates of the translation market in Japan are based on rather slim data.

Johnson, Tim. 1985. *Natural Language Computing: The Commercial Applications.* London: Ovum Ltd. (44 Russell Square, London WC1B 4JP). 459 p. This high-priced publication was intended to brief industry on the state of the art and predict future markets for MT and other NLP software. It was well researched but is now somewhat out-of-date.

[JTEC] Japanese Technology Evaluation Center, Panel Report on Machine Translation in Japan. 1992. Available from NTIS, Springfield, Virginia 22161, as Report PB 92-100239. An extraordinary study by a panel of experts who visited 28 sites in Japan in late 1990. Provides an overview of the state of the art in MT in Japan, and compares Japanese and Western technology.

Kingscott, Geoffrey. 1989. *Applications of Machine Translation: Study for the Commission of the European Communities.* Nottingham (UK): Praetorius Ltd. 81 p. Copyright CEC. The material on history and existing systems is taken from secondary sources and on a number of points perpetuates misinformation that has appeared elsewhere. The report's strength lies in its second half, which examines the translation market, integration of translation with document processing, suitability of texts for MT, translation typology, and possible areas for development. The author is well-informed on the translation business.

Lehrberger, John, and Laurent Bourbeau. 1988. *Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation.* Amsterdam, Philadelphia: John Benjamins. The subtitle of this book reflects its scope. The authors, veterans of the TAUM-AVIATION development effort, offer a protocol for identifying the linguistic characteristics of MT systems, describe the linguistic components and the building of a system, and present a methodology for evaluation which includes such non-linguistic components as the user environment, system maintenance and development, dictionary building, grammar maintenance, specialization of personnel, and the text editor for human revision and documentation of the system. They conclude with global assessment of the system's acceptability. This clearly written volume makes a good introduction to MT for those seriously interested.

Nagao, Makoto. 1989. *Machine Translation: How Far Can It Go?.* Translation from the Japanese by Norman D. Cook. Oxford, New York, etc.: Oxford University Press. While this work starts an overview of MT in general, it quickly settles down to the problems of Japanese-to-English. The author brings lucid examples to illustrate the difficulties inherent in this combination. He concludes with practical proposals for future implementation of MT.

Newton, John. 1991. *Computers in Translation: A Practical Appraisal.* London: Routledge. In press. This book aims to provide an authoritative, readable source on the main areas in which computers can contribute to translation. Of the 12 topical essays, seven are specifically on MT, including papers by Wilks, Melby, Bostad & Vasconcellos, Chandioux, and Somers.

Slocum, Jonathan, ed. 1988. *Machine Translation Systems.* Cambridge, New York, etc.: Cambridge University Press. While this volume contains essentially the same articles and bibliography that appeared in *Computational Linguistics* 11(1-2), 1985, all were updated by the authors in the intervening three years. The contributors are Slocum (survey), Biewer et al. (ASCOF), Vauquois & Boitet (GETA), Bennett & Slocum (METAL), Nagao et al. (Mu), Vasconcellos & Leon (SPANAM/ENGSPAN), Isabelle & Bourbeau (TAUM-AVIATION), and Slocum (bibliography, q.v.).

Vasconcellos, Muriel, ed. 1988. Section III: "The Translator and Machine Translation." In her: *Technology as Translation Strategy.* Binghamton (NY): State University of New York. ATA Monograph *Series* 2. pp. 103-240. Designed as a manual for translators, this hard-cover volume devotes more than half its pages to MT, with clear, highly readable articles on almost all practical aspects of using the technology, by Lawson, Weaver, Smart, Ryan, Santangelo, McElhaney & Vasconcellos, Wheeler, Pigott, Datta, Eng, Newman, Shaefer, Klein, Vasconcellos, Boogaard, and Hutchins.

Zampolli, Antonio, ed. 1989. Special Issue on Machine Translation. *Literary and Linguistic Computing* 4(3). This compendium includes contributions by Landsbergen (Philips), Nirenburg (Carnegie Mellon), Isabelle (Montreal), Maegaard (Eurotra), McCord (IBM), Vasconcellos (PAHO), Rohrer (Stuttgart), Tsujii (Manchester), and Bennett (Texas/ Austin).

Proceedings

The following section lists tutorials, symposia, and conferences that have included MT as a major item on their agenda. In most cases, abstracts or full papers were made available at the time of the meeting. Those proceedings that were later published for wide distribution are listed here with their full bibliographic citation.

[ACL] Annual Meeting of the Association for Computational Linguistics (1962-). The Annual Meeting, held somewhere in the United States each summer, includes numerous theoretical papers on linguistic issues that affect MT, but surprisingly few presentations on MT as such. A European Chapter of the ACL was formed in the early 1980s and also meets each year.

[AGARD] Advisory Group for Aerospace Research and Development. 1990. *Benefits of Computer Assisted Translation to Information Managers and End-Users.* Neuilly-sur-Seine: AGARD, North Atlantic Treaty Organization. AGARD Lecture Series No. 171.

Number 171 in the AGARD Lecture Series was a two-day tutorial on MT geared to potential users, given first in Washington, D.C. (14-15 June), then in Brussels (25-26 June), and finally in London (28-29 June). The presentations had a strong practical orientation and focused on experiences to date. The speakers, whose written papers appear in this volume, were Pigott, Yanez, Gordon, Pinna, Schneider, Bostad, and Lavroff. There is a bibliography (q.v.) at the end.

[Aslib] See "Translating and the Computer."

[ATA] Annual Conference of the American Translators Association.
Each year, at least two sessions are devoted to MT. This conference, attended by about 800 translators, has exhibits of operational MT systems and has served as a forum for presenting new developments and introducing new products on the U.S. market.

[Coling] The program of this biennial conference includes papers on MT.

[Coling'86] 11th International Conference on Computational Linguistics (Bonn, 25-29 August 1986).

[Coling'88] 12th International Conference on Computational Linguistics (Budapest, August 1988).

[Coling'90] 13th International Conference on Computational Linguistics (Helsinki, August 1990).

Commission des Communautes europeenes. 1986. *World SYSTRAN Conference.* Special issue of *Terminologie et Traduction*, 1986(1). Luxembourg: Commission of the European Communities. 202p. Commemorating 10 years of MT at the Commission of Euro pean Communities and the retirement of inventor Peter Toma, the Conference included speeches, reports, and papers by SYSTRAN users and associates around the world, for an audience of some 260 participants.

[GURT'89] Georgetown University Round Table on Languages and Linguistics. 1989. Washington, D.C. Georgetown University Press. Among other scholarly offerings, the volume includes essays on MT by Lawson, Vasconcellos, Zarechnak, Nirenburg, King, Nagao, and Lehmann, marking the 35th anniversary of the Georgetown University Mechanical Translation Project.

[IFTT'89] International Forum on Translation Technology: Harmonizing Human Beings and Computers in Translation (Oiso, Japan, 26-28 April 1989). Program. 64 p. Chaired by Makoto Nagao, this conference brought together more than 400 participants to hear speakers from Japan and elsewhere discussing practical experiences with MT and problems in implementing the technology. The JEIDA Report (q.v.) was introduced at this conference and discussed at length. Speakers from the West included Vasconcellos, Melby, Pigott, Lawson, Rohrer, Wilks, and King.

International Conference on Theoretical and Methodological Issues... See Theoretical and Methodological Issues...

International Workshop on Parsing Technologies (Pittsburgh, 28-31 August 1989).

Kelly, Ian D.K. 1989. *Progress in Machine Translation: Natural Language and Personal Computers,* proceedings of an International Conference on Machine Translation (Cranfield, England, 13-15 February 1984). Wilmslow (U.K.): Sigma Press. This conference was designed to exchange information on the inner workings of MT systems. There were also presentations on the practical use of MT, including a report from the Soviet Union.

King, Margaret, ed. 1987. *Machine Translation Today: The State of the Art. Proceedings of the Third Lugano Tutorial* (Lugano, 2-7 April 1984). Edinburgh: Edinburgh University Press. Information Technology Series 2. 447 p. Although the Lugano Tutorial was held in 1984, this published version of the papers did not appear until 1988, and much had happened in the MT field during the interim. Many of the contributors were associated with Eurotra, which was in its heyday. Some of the introductory papers, e.g., those by Buchmann, Warwick, de Roeck, and Wehrli on the history of MT and Anandiadou's survey, are somewhat selective in their presentation. Sampson ("A Nonconformist's View") and Wheeler (SYSTRAN) are refreshing. Shann gives reasons why a specialized system is not easily extensible to general-purpose translation. System descriptions are given for GETA, METAL, ROSETTA, SUSY, SYSTRAN, and TAUM/AVIATION. The bibliography (q.v.) is very thorough.

Lawson, Veronica, ed. 1982. *Practical Experience of Machine Translation.* Amsterdam, New York: North-Holland. 199 p. The conference by this name, held in London in 1981, gave an excellent overview of MT in actual use, bringing together papers by 19 speakers, including Bostad, Hutchins, King, Knowles, Lawson, Masterman-Braithwaite, Pigott, Sager, Thouin, van Slype, and Wilks. Many of the points made are still valid. This was the only conference in the "Translating and the Computer" series (q.v.) to be devoted exclusively to MT.

[MMT'90] International Symposium on Multilingual Machine Translation '90 (Meiji Kinenkan, Japan, 5-6 November 1990). Program. Held to review progress after four years in the development of interlingual MT among five Asian languages (Japanese, Chinese, Indonesian, Malay, and Thai) by the Center of the International Cooperation for Computerization (CICC), the Symposium brought together representatives from the CICC cooperating countries as well as experts from overseas. The program contains papers and abstracts from 24 presenters, including three from the West.

Machine Translation Summit (Hakone, Japan, 16-18 September 1987). This major conference was the first in a series that brings together representatives from academia, industry, and government who are interested in promoting research, development, and deploy-

ment of MT technology. Abstracts and short papers from 45 presenters are included in the program, which was later updated in a hard-cover version edited by Nagao (1989, q.v.).

MT Summit II (Munich, 16-18 August 1989). Program. 160 p. Following up on the Japanese initiative two years earlier, the University of Stuttgart convened the second conference in this series with assistance from Siemens AG and support from the German government. The program contains abstracts and short papers by some 40 participants.

Nagao, Makoto, ed.-in-chief. 1989. *Machine Translation Summit* (Hakone, Japan, 16-18 September 1987). Tokyo: Ohmsha. 224 p.

Nirenburg, Sergei, ed. 1987. *Machine Translation: Theoretical and Methodological ~.* Proceedings of a conference. Cambridge, London, New York, etc.: Cambridge University Press. International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages. See Theoretical and Methodological Issues in Machine Translation of Natural Languages.

Theoretical and Methodological Issues in Machine Translation of Natural Languages [International Conference on...] The first conference in this series, held at Colgate University (Hamilton, New York, 14-16 August 1985), had partial funding from NSF, and marked NSF's first active involvement in MT since ALPAC (1966). The proceedings were published in a hard-cover book (Nirenburg, 1987, q.v.). The second in the series was held at Carnegie Mellon University and the third at the University of Texas/Austin. It has become a tradition that these meetings alternate with the MT Summits. As the name suggests, the program is always quite theoretical.

Second International Conference (Pittsburgh, 12-14 June 1988).

Third International Conference (Austin, 11-13 June 1990).

Traduction assistee par ordinateur, Seminaire international (Paris, 17-18 March 1989). *Perspectives technologiques. industrielles et economiques envisageables l'horizon 1990.* Paris: DAICADIF. 234 p. This seminar addressed the supply and demand for MT, current markets, and products under development. The compendium includes full-length papers by some 20 presenters (all in French).

Translating and the Computer (London, 1978-). This series of annual conferences, sponsored jointly by the Association for Information Management (Aslib) and the Institute of Translation and Interpreting, has a more practical focus and is attended by some 300 participants from various parts of the language "industry." The program consists of pre-conference tutorials, a "low-tech" day, and a "high-tech" day, on which MT figures prominently. Practical MT systems are exhibited, and sometimes major new products are introduced. Only once was the entire conference devoted to MT (Lawson, 1982, q.v.). The proceedings are published as hard-cover books and widely distributed.

U.S. House of Representatives, Committee on Science and Astronautics, Special Investigating Subcommittee, 86th Congress. 1960. *Hearings...Mechanical Translation Research* (11-13 and 16 May 1960). Washington, D.C. Galvanized by Sputnik and anxious to keep abreast of foreign technology, the United States was investing in more than a dozen MT projects around the country, but early optimism was waning. This was a serious investigation into whether or not research efforts should continue to receive support. It led to appointment of the Automatic Language Processing Advisory Committee, which began its assignment in 1964 and published its results in 1966 (q.v.).

U.S. House of Representatives, Committee on Science, Space, and Technology, Subcommittee on Science, Research, and Technology, 101st Congress. 1990. *Hearing...Status of Machine Translation (MT) Technology* (September 11, 1990). Washington, D.C. 263 p. Aware of the need to keep up-to-date on foreign technology, especially from Japan, and alerted to the potential role of MT by the National Research Council's symposium the year before (q.v.), the above Congressional Subcommittee convened a half-day con sciousness-raising session at which testimony on the potential of MT was given by Harris (NRC), Vasconcellos (PAHO), Wince-Smith (DOC), Brownstein (NSF), Bostad (FASTC/USAF), Carbonell (CMU), Bennett (LRC, University of Texas at Austin), Johnson (IBM), and Zaretsky (Corporate Word). The transcript is presented in this volume, together with appendixes.

U.S. National Research Council. 1990. *Report of a Symposium on Japanese to English Machine Translation* (Washington, D.C., 7 December 1989). Washington, D.C.: National Academy Press. 36p. Interest in keeping up-to-date on technological developments in Japan prompted the National Research Council to look at MT as a means of overcoming the language barrier. In addition to 18 presenters from the U.S., Japan, and Europe, Congressman George E. Brown gave an address that called for planning and long-term commitment to technology development. The symposium was followed by an upsurge of interest in MT among Japan-watchers both in the Government and elsewhere, and served to open the door to a reconsideration of U.S. policy on MT. This report contains a summary of the issues.

Articles of Special Interest

Bedard, Claude. 1988. "You trust your mother, but YOU cut the cards." *Language Technology* 7:26-27, May-June. Six ways that MT demonstrators can stack the deck, and how to avoid being taken!

Bostad, Dale A. 1985. "Soviet Patent Bulletin Processing: A Particular Application of Machine Translation." *Calico Journal* 2(4): 27-30. An interesting application of MT, especially given the current climate, with increasing attention to the tracking of fresh patents.

Bostad, Dale. 1987. "Machine Translation: The USAF Experience." In: *Across the Language Gap: Proceedings of the 28th Annual Conference of the American Translators Association* (Albuquerque, 8-11 October 1987), ed. Karl Kummer. Medford, NJ: Learned Information. pp. 435-443. A description of semi-automated post-editing.

Chandioux, John. 1989. "METEO: 100 Million Words Later." In: *Coming of Age: Proceedings of the 30th Annual Conference of the American Translators Association* (Washington, D.C., 11-15 October 1989), ed. Deanna Lindberg Hammond. Medford, NJ: Learned Information. pp. 449-460. The true story on METEO, which translates weather forecasts in Canada around-the-clock and about which there are some misconceptions.

Joshi, Aravind K. 1991. "Natural Language Processing." In: *Science* (Washington, D.C.: American Association for the Advancement of Science), vol. 253, no. 5025. pp. 1242-1249.

Lawson, Veronica. 1983a. "Machine Translation." In: *The Translator's Handbook*, ed. Catriona Picken. London: Aslib. pp. 81-88. A clear introduction to the field. An update of the *Handbook,* including this article, is in press.

Lawson, Veronica. 1983b. "The Language of Patents: A Typology of Patents with Particular Reference to Machine Translation." *Lebende Sprachen* 28(2):58-61. An interesting experiment on the use of MT to translate patents from German to English.

Lehmann, Winfred P. 1989. "Machine Translation: Achievements, Problems, Promise." In: *Georgetown University Round Table on Languages and Linguistics.* pp. 385-392. An overview of some of the achievements and the prospects for MT, viewed from the standpoint of the work at Georgetown.

Leon, Marjorie, Susana Santangelo, and Muriel Vasconcellos. 1987. "Terminology Work and Automatic Translation Systems: A Case Study at the Pan American Health Organization." *TermNet News* (Vienna) 18:21-25,1987. The difficulties encountered in porting technical terminology from a thesaurus to an MT system.

Loffler-Laurian, Anne-Marie. 1986. "Post-edition rapide et post-edition conventionelle: Deux modalites d'une activite specifique." *Multilingua* 5: 81-88 (Part 1) and 5:225-229 (Part 2). A useful set of criteria for rapid post-editing.

Magnusson-Murray, Ulla. 1985. "Operational Experience of a Machine Translation Service." In: *Tools for the Trade*, ed. Veronica Lawson. pp. 171-180. Pertinent information on the use of MT by a large team of translators.

Marchuk, Yu.N. 1984. "Machine Translation in the USSR." Paper delivered at the International Conference on Machine Translation (Cranfield, England, 13-15 February 1984).

Rudin, Emily B. "Japan's State-of-the-Art in Machine Translation Technology. " *International Science and Technology Insight* 2(3):95-103. A good review of the role that MT can play in capturing information on science and technology in Japan.

Scott, Bernard E. 1990. Letter to the Editor. *Computational Linguistics* 16(4):237-239. An articulate appraisal of Japan's commitment to MT.

Unger, J. Marshall. 1988. "Machine Translation in Japan: Where Are They Coming From? Where Are They Headed?" In: *Languages at Crossroads: Proceedings of the 29th Annual Conference of the American Translators Association* (Seattle, 12-16 October 1988), ed. Deanna Lindberg Hammond. Medford, NJ: Learned Information. pp. 93-102. The author benchmarked raw output from four Japanese-to-English MT systems and two professional human translations.

Vasconcellos, Muriel. 1989. "Long-Term Data for an MT Policy." *Literary and Linguistic Computing* 4(3):203-213. Results of an 11-month experiment to determine, under strictly monitored conditions, whether MT is cost-effective, fast in turnaround, and as serviceable as human translation.

Vasconcellos, Muriel. 1990. "Machine Translation in the 1990s." *Technical Communication* 37(2): 176-179.

Vasconcellos, Muriel. 1991. "Perspectives on the Assessment of Machine-translated Output." In: *FIT Miscellany on Translation Criticism*, ed. Milan Hrala. Amsterdam: John Benjamins. A distinction is drawn between the MT product and the system that generates it, with emphasis on judging the product in terms of the use to which it will be put and on strategies for estimating a system's potential for future performance.

Weaver, Warren. 1949. "Translation." New York. (mimeo) Partially transcribed in Henisz-Dostert, Macdonald & Zarechnak (1979:9-11, q.v.). The famous "Weaver Memorandum." The author, a vice president of the Rockefeller Foundation who at the time was involved in sponsoring research on computers, proposed that natural language translation could be done by computers because language is essentially a code and it maps to a set of universal concepts. The memorandum provided the initial stimulus for research in MT. His points were later challenged, but even so, MT turned out to be feasible.

Wheeler, Peter J. 1983. "The Errant Avocado." *Newsletter of the British Computer Society. Natural Language Translations Specialist Group* 13. The author, inspired by a SYSTRAN mistranslation, develops a clear explanation of ways in which the system's dictionaries can be manipulated to obtain context-sensitive translations.

Copies of this report may be obtained
from the National Technical Information Service
5285 Port Royal Road, Springfield, Virginia 22161

Phone 703-487-4650

Request # PB93-134336