APPLICATIONS OF MACHINE TRANSLATION

STUDY FOR THE COMMISSION OF THE
EUROPEAN COMMUNITIES

Geoffrey Kingscott
Praetorius Limited

Produced by Praetorius Limited, 5 East Circus Street. GB - Nottingham, NG1 5AH

For the Commission of the European Communities.

September 1989

APPLICATIONS OF MACHINE TRANSLATION

TABLE OF CONTENTS

Table of contents (continued) **Page**

Table of contents (continued) **Page**

SECTION 1 INTRODUCTION

This report looks at the progress made in machine translation up to the date of writing, the likely future development, the context in which such development will take place, and the economic and cultural implications of such development. The emphasis throughout is on applications, and no attempt is made to provide a technical description of the linguistic features of the various machine translation systems mentioned.

We begin with a definition of terms, since sometimes terms such as machine-aided translation are used loosely, and distinctions of meaning are lost. This is followed by an examination of the context in which translation takes place, looking at the current demand and the potential demand for translation, and how this might be met by machine translation. The need for a survey of the translation market, as a basis for rational decision-making in this field, is explained.

The early history of machine translation is recalled, and the fact that progress in machine translation is often a direct response to a perceived need is examined. There then follows a major section which surveys the world's operational machine translation systems, while a shorter section which follows takes a brief look at some of the systems still in the research and development stage, and consideration of the market share held by machine translation companies.

The documentation available on machine translation is examined in section 10, and this is followed by a typology of translation, showing the suitability or non-suitability for machine translation of different categories. The important question of the integration of translation into document processing generally is the subject of a short section, and another short section indicates the diversity of approaches. The applicability of machine translation to different types of translation activity is investigated in section 15, and section 16 looks at possible areas of developing work associated with machine translation.

Finally suggestions are put forward for better promotion of machine translation.

SECTION 2 DEFINITION OF TERMS

A distinction is usually drawn between:

machine translation, meaning automatic translation of a whole document done by computer, as a general rule carried out in batch mode;

machine-aided translation, normally taken to mean translation where the computer produces the translation in short sections for the translator or post-editor to work on, as a general rule carried out in interactive mode; and

machine aids to translation, where the translation is produced and input directly by human translators; such aids include word processing programs specially aimed at translators; ability to access foreign language character sets, accessing of remote databases by modem; machine-readable glossaries and dictionaries which can be brought up on screen during the translation process; spelling, style and grammar checkers, etc.

Machine translation is sometimes called computer translation or automatic translation, and machine-aided translation is sometimes called computer-aided translation. The abbreviated forms MT, MAT, CT, CAT are also found.

Machine translation does not exclude human intervention, and this is normally necessary - at the present state of the art - in the form of either pre-editing of the source text, or post-editing of the target text, or both.

Unedited machine translation is generally referred to as raw translation. The word raw in English has connotations of "crude", "unrefined", "unfinished". It is possible that these connotations have inhibited the use of raw translation, since there are purposes for which it might be suitable.

Human translation is the term usually used where the actual translation is generated by the human translator, whatever the tools used, pen and paper, dictation, typing, or inputting into a computer using a word processing program. It does not exclude the use of machine aids to translation.

The traditional distinction between Human-assisted machine translation, and Machine-assisted human translation, once quite clear, is becoming somewhat blurred.

However, for the purposes of this study of "applications of machine translation" it is proposed that "machine translation" be widely interpreted, to comprehend all translation applications where target language text is automatically generated, as a whole, or to a substantial extent.

SECTION 3 THE DEMAND FOR TRANSLATION

It is important to place the issue of machine translation in the context of actual and potential demand for translation work in the world. If, as many authorities assert, demand is increasing faster than the pool of experienced human translators available to do the work, then it may be necessary to devote more effort and investment into developing machine translation.

Unfortunately there is a lack of reliable information in this area.

The following figures have been quoted at different times by various authorities, though the basis for the figures is rarely given.

An estimate of the world translation market, given by Mr Loll Rolling of DG XIII of the Commission of the European Communities, speaking in April 1989 at the IFTT conference in Tokyo, is 200 million pages a year, "which only reflects the cash flow between client companies and translation bureaux manned with literary translators. It does not cover the huge amount of translation work that is performed within the companies".

Another estimate is 18 to 23 million pages a year in Japan. An estimate made in 1986 by the Japan Electronics Industry Development Association was that the total translation work amounted to US$ 5.5 billion (approximately 5 billion ECU), mostly between Japanese and English.

Another estimate, the Dataquest survey 1986, is $500 million (455 million ECU) for machine translation.

Another estimate is that in 1986 more than 100 million pages were translated in the West European countries (Alonso and Schneider).

At a hearing on the Language Industry in the European Community held in Luxembourg in December 1988 it was estimated that in the EC alone some 100 million pages of text are translated a year, corresponding to a work force of 100,000 and a turnover of 10 billion ECU.

The Translation Bureau of the Canadian Secretary of State has calculated that the number of words a year which it translates is about 300 million (approximately 120,000 pages), of which 80 per cent is between English and French.

Another estimate, given in an article by Ian Pigott, *Operational Machine Translation Systems,* is that the requirement for translation is growing by an average of about 15% per annum.

Juan A. Alonso and Thomas Schneider, Machine translation technology: on the way to market introduction, in Siemens Review, vol. 54, no. 6, November/December 1987.

APPLICATIONS OF MACHINE TRANSLATION

One estimate of the growth in machine translation which does give the basis of calculation was made by computer systems consultant Tim Johnson in 1985, in his report *Natural Language Computing: The Commercial Applications,* published by Ovum Limited, London. He produced projections for the growth in the use of machine translation systems. He forecast moderate growth to 1989, "until a significant proportion of about 10,000 professional translators in the USA are using them; then more rapid growth as machine translation becomes an essential aid to the best quality technical translation, and improvements in cost-effectiveness make it a useful tool for information scanning and non-professional translators, creating new markets for translation". The forecast was evidently far too optimistic about the speed of development, but the reasoning about the pattern of development is probably sound.

One problem in estimating the size of the translation market is that professional translators - the mainstream of translation activity - probably account for less than half of translation work carried out in the world, especially in the private sector. Indeed in the opinion of the author of this report the figure may be as low as 20%.

A comparatively recent factor is that authors of scientific papers and technical manuals are now frequently writing these in a language not their own, usually in English; this is not translation activity as such, but they would undoubtedly benefit if they were able to produce such work through the medium of machine translation.

Further, machine translation researcher Sami Trabulsi, speaking at the conference on *Problems of professional translation in the Arab world and the contribution of new technology in Tangiers* in June 1989, estimated that only 2.5% of all documentation produced was translated into one other language, let alone all the languages where translation could be useful. To give but one example, students of subjects such as accountancy or car mechanics in countries such as Malaysia, Thailand or the Gulf Arabic states often have to use English textbooks, despite a limited command of the English language, because of the dearth of material in their own language.

SECTION 4 SURVEY OF THE SIZE OF THE TRANSLATION MARKET

What is badly needed is an objective assessment, based on hard statistical evidence, of the size of the translation market, from three points of view:

the amount of material at present being translated, and the types of material involved (present market);

the likely increase in the type of material at present being translated (future market);

the amount of material which it would be useful and cost-effective to translate if there were improvements in the speed at which translation is performed, or there was much more widespread accessibility to translation resources, or if costs to the translation user could be lowered (potential market).

Appendix I sets out proposals for the carrying out of a survey of the size of the translation market.

SECTION 5 EARLY HISTORY OF MACHINE TRANSLATION

The history of machine translation has been told before, most notably as far as works in English are concerned, in a book by W. John Hutchins, of the University of East Anglia, UK. A survey of systems, and a useful bibliography, can also be found in a more recent book edited by Jonathan Slocum. Mr Hutchins, who endeavours to keep abreast of all machine translation developments, and whose assistance in the provision of information for this report is gratefully acknowledged, has brought his history up to date with a paper given at a conference in Budapest in 1988 and with a paper to be presented at a conference on machine translation in Tbilisi, Georgia, in the Soviet Union, in November 1989, of which an advance copy was made available to the author of this report.

There are of course new books, journal articles and conference reports appearing all the time, since this is a developing field. Section 10, Description of Experiences, looks at these in more detail.

It is obvious to all who write on the subject that the history of machine translation falls into distinct phases. Speaking at the 1981 Translating and the Computer conference Mr Hutchins identified four phases, and his division and suggested dates were echoed in a Memoir of 1983 by Chantal D'Hondt. In this she suggests:

1946-1954, Premières Expériences;

1954-1966, Optimisme et Désenchantement;

1966-1975, Diversification des stratégies; and, depuis 1975, Renouveau d'optimisme.

It can of course be argued that the impact of artificial intelligence in the 1980s marks another stage, but this report is concerned with the applications of machine translation, and essentially in recent years no huge advances, in the sense of marking a new era, have been made on the applications side.

Indeed older systems still rule the roost, and even if METAL promises to be a serious competitor, its applications are still limited to a few sites which are still accustoming themselves to its use.

W. John Hutchins, Machine Translation: Past, Present, Future, published by Ellis Horwood Ltd, Chichester, UK, 1986.

Jonathan Slocum (ed.), Machine translation systems, published by the Cambridge University Press, 1988.

Recent developments in machine translation, published in New directions in machine translation, edited by Dan Maxwell, Klaus Schubert and Toon Witkam, published 1988 by Foris, Netherlands.

Chantal D'Hondt, Systran et la Communauté Européenne à Luxembourg. Contribution au Dossier "Traduction Automatique", Mémoire de fin d'études, EUR 9747 FR, 1985.

APPLICATIONS OF MACHINE TRANSLATION

Since the computer was first developed (essentially, if precursors are to be ignored, in the 1940s) researchers have had the idea of using it for the translation of natural language text. One early suggestion, by pioneer Warren Weaver, was that language was merely a form of code, and as much of the early development of computers was connected with the breaking of enemy codes during the second world war, the application seemed a natural one. Although code-breaking was never actually adopted as a method, in the early years there was a widespread notion that translation was essentially a matter of word substitution using a bilingual dictionary.

If a sweeping generalisation can be permitted, the history of machine translation can be said to consist of a growing but only gradual realisation of the almost infinite complexity of natural language, and the difficulties inherent in processing it. As Lehrberger and Bourbeau comment, "Mechanical translation is perhaps the first attempt to apply computers to the simulation of a non-numerical human activity".

The word substitution notion quickly gave way to the writing of paradigms to allow parsing and semantic analysis to take place, and early systems began to emerge.

One of the leading early systems which found practical application was the Georgetown University project in the United States. To summarise in very broad terms what was quite a complex history, Georgetown University, with government support, started work on its GAT project in 1952 for the translation of Russian physics texts into English. The system became operational in 1964 with delivery to the Atomic Energy Commission at Oak Ridge National Laboratory, and to the corresponding research facility at EURATOM in Italy.

Despite what by all reports was a crude output, it is understood that this original system, or something very close to it, continued to be used for information scanning of Russian documents at EURATOM until 1976 and at Oak Ridge to about 1979. As Jonathan Slocum *(op cit)* comments, "The fact that it was used for so long is nothing short of remarkable - a lesson in what can be tolerated by users who desperately need translation services for which there is no viable alternative to even low-quality MT" (EURATOM finally acquired Systran as a replacement for GAT in 1976).

However, in circumstances that are still obscure, the Georgetown University system ceased to receive funds in 1963.

About this time a change of mood became noticeable, which was to result in a widespread feeling that machine translation was not getting anywhere, culminating in the ALPAC report of 1966.

ALPAC is an acronym for the Automatic Language Processing Advisory Committee of the United States Academy of Sciences, and the committee had been established in April 1964, to study the existing demand, supply and costs of translations. It concluded that the supply of human translators exceeded the demand, and that machine translation had progressed from the "deceptively encouraging" early results to "uniformly discouraging" results at the time of the report.

John Lehrberger and Laurent Bourbeau, Machine Translation - Linguistic characteristics of MT systems and general methodology of evaluation, John Benjamins, Amsterdam, 1988.

It is true that expectations in the 1950s and 1960s, that Fully Automatic High Quality Translation (FAHQT) of non-selected texts might be possible with the computing knowledge of the time, were unrealistic, but the National Academy of Science's experts were also influenced by a mood of disillusionment which had set in among some of the early researchers.

But the ALPAC report's effect on United States government support for machine translation development was devastating, and by 1973 only three government-funded projects were left in the United States, and by late 1975 machine translation research in the United States had reached its lowest point.

Machine translation research had not totally disappeared, and three projects that were still going on at the end of 1975 have been identified. One was research at Brigham Young University which, although unknown to anyone outside Provo, Utah, was to result in the ALPS and Weidner systems (see below); another was the Xonics system (Russian to English), which was demonstrated at the FBIS seminar in 1976, and which was a precursor of the SPANAM prototype, and Systran, which, it was later claimed, had been used in the Apollo-Soyuz project of 1974 to 1975.

Nevertheless it is true to say that the effects of the ALPAC report, on funding for machine translation research, and on attitudes and expectations, can hardly be exaggerated. There were frequent references to it at the two major machine translation conferences so far held in 1989, the IFTT conference in Tokyo in April, and the Machine Translation Summit II held in Munich in August.

One result of the ALPAC report was that machine translation research became a peripheral activity in the United States, carried on only by maverick organisations (such as the Mormon church, which sponsored much of the work at Brigham Young University) or by maverick individuals, such as Dr Peter Toma, developer of Systran, maverick not in any disparaging sense, but in the sense that they went their own way. Even today many of the companies selling machine translation systems in the United States are not major industrial firms. This should be contrasted with the current situation in Japan, where machine translation activity is being carried on by companies such as Fujitsu or Mitsubishi, the giants, that is to say, of Japanese industry.

The history of machine translation development after ALPAC was to a large measure the history of individual projects, and these are dealt with in section 7, below.

Language and machines: computers in translation and linguistics, A report by the Automatic Language Processing Advisory Committee, Division of Behavorial Sciences, National Academy of Sciences, USA, 1986.

SECTION 6 PERCEIVED NEED

Although the ALPAC report was a major setback to development, particularly at government level, the perceived need for automation of the translation process meant that some projects continued, and the first commercially available systems did emerge.

Perceived need has changed with time, place and circumstances, but is often a response to particular conditions.

It is perhaps forgotten now just what a psychological impact the launch of the first Soviet Sputnik spacecraft on the 4th October 1957 made on the complacent West. It was a revelation that Soviet technology was so far advanced that it was the Soviet Union, and not the United States as everyone expected, which had succeeded in putting the first man into space.

Suddenly there was a realisation that in the West we needed to peruse Soviet scientific literature, and this had led to considerable US funding for machine translation in the late 1950s. The impact of Sputnik is clearly reflected in the concentration on systems to translate from Russian to English; in the mid-1960s both the Georgetown University system and the IBM Mark II were used for military and political purposes. The United States Air Force application of Systran (see below), after 1970, took over from the IBM Mark II and so was a continuation of this policy.

Another "perceived need" arose out of developments in Canadian society in the 1970s, in particular with the passing of the first Official Languages Act (1969), which made Canadians realise the huge volume of translation work this implied. It should not be forgotten too, that in the dark post-ALPAC days of the 1970s in the United States, the Canadian government supported for eight years the TAUM (Traduction Automatique Universite de Montreal) project at the University of Montreal, one of the results of which was TAUM-METEO, one of the first machine translation systems to work cost-effectively.

The growth of Japanese trade with the West, which began on a large scale in the 1970s and reached gigantic proportions in the 1980s, stimulated machine translation work in Japan and attracted government backing in various forms.

And the geographical widening of the European Community, and also, and perhaps more importantly, the enlargement of its spheres of activity, has made the automation of translation between the languages of all member countries, but in particular the main languages, an imperative.

In practice, it appears, in the early stages of European unification, French was something of a lingua franca. The admission of the United Kingdom meant the admission of another major world language, and there was no longer a single lingua franca. Translation between English and French to begin with, and then between those languages and other Community languages, and then between the other languages, has been the order of priority determined by the political circumstances.

What other language imperatives can be foreseen, apart from the accelerating growth of international trade, and a concomitant growth in documentation for products which become ever more complex?

One such imperative may emerge from the political will to create a Francophonie. The term "language industries", now widely used, was coined in this context. In the words of Alain Landry, Canadian Assistant Under Secretary of States, "Cette nouvelle appellation (les industries de la langue), issue du premier sommet des pays francophones, qui a eu lieu à Paris en 1986, sert à souligner à la fois les structures qui se sont développées dans les pays industrialisés au cours des derniers vingt ans et l'apport qu'elles sont appelées à fournir aux différents projets d'aide et de développement que la francophonie a lancés." We may then see a much more sustained effort than hitherto to ensure that necessary information is translated into French, not just for the developed countries such as France, Belgium, Switzerland and Canada, but also for the Third World French-speaking countries, and for a resurgence in the use of French in resistance to the ubiquitous use of English.

A similar development may happen in the Arabic-speaking countries. Divided as they may be in political questions, they are united by a common language of which they are proud, and are now actively seeking greater co-operation on language questions, as evidenced by two conferences held this year, on the coordination of translation activity (Tangiers, June 1989) and on the standardisation of terminology (Tunis, March 1989).

There seems little doubt that sensitivity to language questions, and pride in one's own language, is growing in countries such as Portugal, whose first language fair, in September 1989, attracted more government support than any of the other language fairs which are now held in many European countries. At present translation resources in Portugal are scanty, but the demand for translation into Portuguese is growing rapidly.

The policy of *perestroika* in the Soviet Union may generate huge new demands for translation. Up to the beginning of this year Russian was the only language used in international trade and relations with the outside world, but we are now likely to see demands for more documentation to be provided in languages such as Estonian, Latvian, Lithuanian or Moldavian. And if economic liberalisation gains ground in the Eastern block countries generally, languages such as Hungarian or Polish may become economically important, despite the lack of translators between these languages and, say, English or French.

Alain Landry, Opening Address, First Congress of the Canadian Translators and Interpreters Council, Ottawa 1987, proceedings published by Linguatech, Montreal.

SECTION 7 SYSTEMS IN CURRENT USE

In John Hutchins's history of machine translation some 60 initiatives in developing systems are recorded, but in this section of this report it is the intention to concentrate on the widely-known systems in current application.

7.1 TAUM-METEO

Quite a few applications have been found for the 1970s technology. One example, often quoted, is the METEO system which has been used in Canada for about ten years, and since 1977 has translated about 15 million words per year from English to French without significant human intervention. This year, 1989, it has started on translating from French into English.

The full title is TAUM-METEO, since it resulted from work done under the TAUM *(Traduction Automatique de l'Université de Montréal)* project at the University of Montreal, where the prototype was produced. The TAUM project was commenced, with Canadian government funding, in 1965. According to some reports, early research was fairly unstructured, and it was the Canadian government sponsors who suggested that the development work be put to specific projects.

TAUM-METEO was commissioned in 1975, the prototype demonstrated in 1976, and the production model put into operation in June 1977. The six-year federally-funded research effort had been initiated in the early 1970s by the Canadian National Research Council and was later taken over by the Secretary of State's translation bureau.

Improvements led to METEO2 in 1983. With a relatively small dictionary of some 2,000 words and expressions, the system translates hourly weather forecasts for Environment Canada's meteorological stations in Vancouver, Edmonton, Calgary, Regina, Toronto and Halifax. This year (1989) saw the launch of the new French to English version in Montreal, which is now providing computer translation of all Quebec weather forecasts.

METEO handles an average of 30,000 words a day, or 8.5 million words a year, and appears to work to everyone's satisfaction. Not only is METEO one of the longest-running machine translation systems in practical use, but it is also, of course, a classic example of the suitability of machine translation for specific-task and repetitive-task situations. Indeed before the installation of TAUM-METEO, one of the problems of the Canadian Meteorological Centre was the high turnover of translation staff - the average length of stay was six months - which was attributed to the monotony of their work (since the introduction of TAUM-METEO this problem has disappeared).

There must be many more task-specific situations which would benefit from current state-of-the-art MT, and this point is explored further in section 18, promotional activities.

It is a matter of record that a second project, TAUM-AVIATION, designed for the translation from English into French of a specific set of aviation maintenance manuals, totalling 90 million words of source text, had to be abandoned because of spiralling costs, but this was some ten years ago. There is no reason to consider the task-specific approach has thereby been invalidated, and recent developments, such as the CALLIOPE-AERO and CALLIOPE-INFO research projects in France, and the customisation by Lexi-tech in Canada of first Logos and then Tovna for the translation of shipbuilding and ship maintenance documentation, show that this is still one of the avenues for MT progress.

7.2 Systran

One of those who had worked on the Georgetown machine translation project (see above) was Dr Peter Toma, who disagreed with the conclusions of the ALPAC report and insisted on persevering with machine translation development. He created two systems, Autotran and Technotran, which were forerunners of Systran. He described his colourful early experiences (he had a day job between 8 and 4.30 p.m. slept between 6 and 10.30 p.m., and worked all night - the only time he could get uninterrupted access to the computer - on machine translation algorithms) at the World Systran conference in 1986. By 1970 he had a Systran system which he was able to deliver to the United States Air Force Foreign Technology division at Wright Patterson Air Force Base, where Systran is still being used today for information scanning of scientific and military information in Russian, and there seems to be no reason why it should not celebrate 20 years of operation next year. The current throughput is some 100,000 pages a year. French to English and German to English versions are now being introduced to the system by the Air Force, but as yet there seems to be no report published of their operation.

Systran was first acquired for development by the Commission of the European Communities in the mid 1970s (the system was demonstrated in 1975, and their contract signed in February 1976), and currently a team of 35 people, either employed by the Commission or contracted to it, are working on the development of 12 language pairs, English into French, Italian, German, Dutch, Spanish and Portuguese; French into English, German, Dutch and Italian; and German into English and French. German is notoriously a difficult language for machine translation, and it was not until 1987 that development work really got under way on German as a source language.

The 1976 acquisition by the Commission represented a major step forward for the system, since up to that time Systran had been seen as essentially an information-scanning system. The European Commission, however, were interested in the translation of texts for distribution, which meant that texts had to be produced, either by the system or by post-editing, to a much higher standard.

According to Ian Pigott, head of the Commission's Systran project, the Commission's versions of Systran have become steadily more modular over the years, and source language analysis is now completely independent of target language generation, even in respect of the basic dictionary. Dictionaries have been enlarged and this and the greater modularity have made the system gradually more usable.

Dr Peter Toma, Systran's contribution to mankind, in Proceedings of the World Systran conference, Luxembourg, February 1986, published by the Commission of the European Communities.

Translations are processed at a rate of about 500,000 words an hour, and now the system is available on Unix there are over 400 user work stations. It may be that making the system available on Unix may prove to have been a major step to accessibility.

The European Commission has certain user rights over the system in connection with its use by organisations in the public sector in the EC member states, and works closely with those organisations who use Systran, including NATO headquarters in Brussels, the Nuclear Research Centre in Karlsruhe, and German Railways.

The development work at the Commission has been carried out by as a joint action by Directorate-General XIII (Telecommunications and Information), and Directorate-General IX (Personnel, including translation, converted into a separate Directorate-General in 1989).

The first commercial uses of Systran date back to around 1979, though apparently General Motors of Canada had acquired a version as early as 1976. Other long-term users in the commercial field include the Xerox Corporation, which uses Systran in conjunction with a controlled English input, for technical manuals, out of English into six target languages, Dornier and Festo, working from German into English, and a translation company in Tokyo which uses English to Japanese and Japanese to English versions (there is a Systran Japan company).

World rights over Systran in the private sector are now owned by the French company Gachot SA, of Soisy-sous-Montmorency, near Paris, which has an agreement with the Commission for coordination of work. The Gachot company, in an unusual step, has made Systran translation available on the French Minitel network, with its three million subscribers, since the beginning of 1987.

Gachot is currently offering 12 language pairs, English to Arabic, French, German, Italian, Spanish and Dutch; French to English, Dutch and German; German to English and French, and Russian to English. The speed of translation is said to be 500,000 words an hour, or 26 pages a minute. There is a central computer at the company's headquarters. Gachot employs some 100 people in Europe and in the United States on developing the system.

There is no doubt that one of the reasons why Systran has held its own at the head of the machine translation field for so many years is the increasing amount of feedback received from users. The world Systran conference in February 1986 was a major event in this respect. Another factor is the richness of its dictionaries, built up over so many years. The system is sometimes criticised as an "old" system, but in fact has been so modified and adapted over the years that it can be regarded as a current system.

Ian M. Pigott, Systran at the Commission of the European Communities, paper given at the conference IFTT '89, Tokyo. April 1989.]

7.3 SPANAM and ENGSPAN

SPANAM and ENGSPAN are used at the Pan American Health Organisation (PAHO) in Washington DC, and at some other sites. It is now the principal mode of translation between the organisation's two working languages, English and Spanish. In the person of its director, Dr Muriel Vasconcellos, and in the early development of SPANAM, the systems date back to the early days of research into machine translation at Georgetown University, Washington. Work on SPANAM, the Spanish to English system, was begun in 1976 and was ready for implementation in late 1979, and an updated version came into use in November 1988; ENGSPAN, constructed on somewhat different principles, has been in use since 1984.

The SPANAM system was produced by outside consultants, but in 1980 the source code was delivered to PAHO, and since then responsibility for the continued development and operation of the system has lain with the PAHO translation staff. Average output is said to be 6,500 words a day per translator cum post-editor (the usual range is said to be 4,000 to 10,000 words a day, which is between two and three times as fast as the previous conventional translation output of 1,500 to 3,000 words a day).

Work, it has been announced, has commenced on the development of an English to Portuguese version (ENGPORT), but there is no recent information on the latter.

One of the few instances of measurement of quality of translations processed using machine translation, as measured by customer response, has been carried out by the Pan American Health Organisation. This found slightly higher percentages of respondents were more satisfied with the post-edited machine translation output than with the translation work entirely done by human translation.

Muriel Vasconcellos, MT Utilisation at the Pan American Health Organisation, paper given at the conference IFTT '89, Tokyo, April 1989.

7.4 TITUS

TITUS, a system developed for the Institut du Textile de France, is an example of a text-specific system used in restricted circumstances, for translation between French, English, German and Spanish of abstracts of articles on various aspects of textiles. TITUS is unusual in that it requires rigorous pre-editing, so that sentences, for example, are always written in the same order, Subject Verb Complement, while such problem items as relative pronouns etc. were banned in the earlier versions.

The system first came into use in the 1970s, and has undergone some evolution. The version TITUS 4 was implemented in 1980 and allows the introduction of certain subordinate clauses, relative clauses and "that" clauses, thus allowing a closer approach to a capacity to translate free-language sentences or phrases. The basic ideas, however, of restricting language to the subset of textile vocabulary (apart from necessary general vocabulary) and to a subset of the syntactic rules of the working languages, remains.

Because of the need that all sentences of a record to meet the TITUS 4 syntactic rules and pre-established vocabulary, certain computer routines have been developed to check input for errors or ambiguities, such as polysemy or homography. These routines display questions or messages on screen, and thus TITUS 4 can be said to have developed into an interactive system.

The languages in which TITUS works are French, English, German and Spanish.

Because users of the system must comply with the writing rules imposed by the controlled syntax, training is required, and the time taken for this is estimated at between five and six days.

TITUS 4 is an example of a task-specific system. It can also be regarded as an example of a transfer system, since each sentence is transformed into what the developers call a "swivel language" before storage. The "swivel language" is a binary form language whose particular structure permits very quick translation of a sentence in one or more of the four languages.

A.A Strerff, New developments in TITUS 4. in Veronica Lawson (ed.), Tools for the Trade, proceedings of the Translating and the Computer 5th conference 1963, published by Aslib, London.

7.5 Weidner

The Weidner company was established in 1977 in the United States by Bruce Weidner, who took the opportunity of the winding down of the machine translation research project at Brigham Young University, Utah (see above) to hire some of the programmers.

The first commercial user we have been able to trace was Mitel, who took delivery of an English to French system in 1980, and subsequently acquired further systems for translation from English to Spanish and English to German. Marketing in Europe commenced about the same time -the author of this report recalls attending the first demonstration in the United Kingdom - and a system was acquired by ITT at Harlow in Essex, which under its first manager, Ulla Magnusson-Murry, turned into an important pioneering venture for MT.

Weidner is available on MicroCAT (personal computers) and MacroCAT (mini-computers); language pairs available include English to French, Spanish, German, Portuguese, Italian; French to English; Spanish to English, and, on MacroCAT only, English to Arabic and German to English. Speed of raw translation is said to be about 4,000 to 8,000 words an hour or between 16 and 32 pages. It is claimed that a trained post-editor can produce 600 to 1000 words an hour.

TAO International market Weidner in France, and also act as a translation office, using the system for their own purposes (a description of how they use Weidner was given at the conference *Journées européennes de la traduction* in Paris in 1987). Other users in France include or have included Aerospatiale, Bull, Matra (through SAC Control), Motorola, Télésystemes, and Thomson in France, and in England Perkins Engines, Rank Xerox and Massey Fergusson.

The Perkins Engines experience was described at the 1988 Translating and the Computer conference (proceedings not yet published) and again at the London Language Show in June 1989 by the head of technical publications, Peter Pym. The key to their successful operation of the system is the writing of the original documents in controlled English, which they call PACE (Perkins Approved Clear English).

Aérospatiale, who have probably as much experience with Weidner as anyone apart from Mitel and ITT, are said to have achieved a 50% increase in productivity.

Weidner, whose marketing company by this time had become known as World Communications Corporation, began working with the University of Bergen on a system for translation English to Norwegian, one of the first MT projects to encompass a Scandinavian language, but the project never reached the stage of actual application and now seems to be in abeyance.

One of Weidner's first reference sites in Western Europe, as stated, was part of the Engineering Support Centre of ITT, at Harlow, Essex, UK. When ITT was taken over by Alcatel, it was decided to dose the Centre, but Centre executives decided on a management buy-out, and set up a new translation bureau company, ESC.

Heinke Koppen, Le système Macrocat (Weidner), proceedings of the conference Journées européennes de la traduction professionnelle, published by Encrages, University of Paris VIII, 1987.

A question mark hangs over the future of Weidner at the present time. The former company which operated out of the United States, World Communications Corporation, was acquired by Bravice International Inc., a Tokyo company, effectively owned by Takehito Yamamoto. In 1988 Mr Yamamoto closed down the United States and European operations, and has not yet set up replacement organisations for technical back-up and marketing.

In the United States there may be some legal problems about rights of ownership of software codes.

Certainly Bravice now seem to be concentrating their efforts on the Japan side of their operations, where they have been reasonably successful in selling systems, and in new developments - such as new Toolkit system, not yet released - rather than in the existing applications.

7.6 ALPS

Like Weidner, ALPS can trace its ancestry back to a project, established at Brigham Young University in 1971, to translate texts, both sacred and promotional, for the Mormon Church out of English into many other languages, starting with French, German, Portuguese and Spanish. When the project was discontinued, one group joined the Weidner team, and the other formed Automated Language Processing Systems (ALPS) in 1980.

The ALPS system is unusual, in that it can function on three different levels, the higher levels encompassing all the features of the lower level or levels, and that the highest level, which provides automatic translation, is interactive, that is it offers a proposed translation to the translator sentence by sentence, with questions to elucidate ambiguities.

The three levels, in rising order, are Selecterm, Autoterm and Transactive, and the languages pairs offered are: English to French, German, Spanish, Italian; French to English; German to English. Users include Texas Instruments, Unisys, Coventry Polytechnic, NCR France, NATO, Norsk Data, Union Bank of Switzerland, Rank Xerox Limited, and L & L Gebruikers Informatie, Netherlands, as well as translation companies which are now in the Alpnet international group.

This group was created in 1987 to 1988, as a result of a deliberate switch of priority by ALPS, from selling software systems to operating a translation bureau service. The three largest businesses of their type in the United Kingdom (Interiingua-TTI), Federal Republic of Germany (Haehl) and Canada (Multiscript-La Langagerie) were acquired.

There is evidence that some purchasers of the ALPS system, after a period of experimentation, often do not trouble to use the highest level, Transactive, but do find the second level, Autoterm, a distinctively useful translation tool.

Against this, however, it should be stated that there are apparently satisfied users of ALPS Transactive mode, at NATO Brussels, for example, and one translator at Statistique Canada in Ottawa. The answer may be that to use Transactive successfully takes a great deal of perseverance and time, and that in the translation bureau situation there is little opportunity to spend enough time mastering the art of using the Transactive mode, but with encouragement, in the less stressed public services, Transactive can be made into an effective system.

At Multiscript-La Langagerie in Montreal Shiera O'Brien, Customer Support Manager for TSS (Translation Support System), confirmed they used the Autoterm level, and said that a text had to be over 30,000 words to make it cost-effective to put through the system. She calculated an increase in output of seven to 15% was achieved with the system, even taking into account the need for dictionary build-up. It took the average translator five days to get used to the machine. The text for translation by this method should have a high technical content, be repetitive, and be written in a reasonably simple style.

This experience seems to be confirmed in many of the same terms in a discussion by translators who have worked with the system at the London offices of Interlingua Limited, another part of the ALPNET group.

7.7 Logos

Logos is another American company, and the actual system development began in the 1960s. A version for translation from English to Vietnamese was introduced in 1971 for the translation of military equipment maintenance manuals. After a false start, Logos produced at the start of the 1980s a German to English system which ran on Wang computers.

This was at a time when system designers were a little wary of German, notoriously one of the more difficult languages from a machine translation point of view.

"Arriving on the MT scene within the past couple of years as an apparent newcomer (despite its history stretching back to 1969), Logos has already established itself as the one to beat in German-source MT" (Peter Wheeler at the 1984 Translating and the Computer conference).

Language pairs now offered are German to English and French. English to French, German and Spanish, and it is claimed that the system will translate, in batch mode, up to 100 pages an hour. Users in Europe include Nixdorf, IBM Germany and Hewlett Packard. All Logos systems are now available to run on VM/CMS mainframes in addition to the traditional Wang office systems versions.

In Canada evaluation of Logos is being carried out for the Secretary of State, under the guidance of Alain Landry, Assistant Under Secretary of State, and on May 16, 1989, the bureau opened its *Centre d'*expertise on Computer-assisted translation. This is "to co-ordinate the activities of the various CAT sites, and to conduct an evaluation of related technologies" on the subject.

According to the official announcement. "The establishment and operation of the centre were made possible thanks to the cooperation of Wang Canada Limited, who offered the equipment (VS-5000 with software and peripherals) with maintenance and technical assistance, and Logos Canada Limited, who offered the software licence without charge for the duration of the operations of the centre."

Peter Wheeler, Update on machine translation, in Translation and communication, ed. Catriona Picken, Proceedings of the sixth Translating and the Computer conference, November 1984, published by Aslib, London.

7.8 Smart

A system which has had some success in North America, but which to date has not been marketed in Europe, is the Smart system of Smart Communications Inc. of New York. The company, which also markets a system for writing manuals in clear English, produces the Smart Translator which, it is claimed, can produce raw translations at a speed of 200,000 words an hour when running on mainframes.

Smart's largest customer is the Canadian Ministry of Employment and Immigration in Ottawa. The system is used to translate job descriptions between English to French, and these can be displayed within minutes at any of the Department's 3,500 terminals installed in its offices throughout Canada. The translations are generated on screen, and quickly post-edited into acceptable French by the operator.

Complaints were quite rare, and usually only from pedantic teachers complaining about grammatical solecisms, and hardly ever because the description was misleading. It was quite dear, however, that the system worked, and worked quickly, and was being widely used virtually without complaint. As more and more job description idioms were fed into the system, the generated text was becoming more accurate. There was no attempt to control the language of the source text, except that there was a limitation on length, and of course such texts, supplied by employers, varied considerably in style.

It will be apparent that this task-specific use of machine translation could have an application in Europe. With the advent of the Single European Market in 1993, and with the already established free movement of labour within the Community, it would be an advantage to have at least some job descriptions more widely circulated.

The author of this report raised with the Canadian employment officials the obvious point, that only those people who are already bilingual would think of moving out of their own language area, and such people therefore did not need bilingual job advertisements. I was told, however, that while this might be true of the professional classes, there were some classes of worker such as welders and steeplejacks who were in great demand, and the salaries offered could well attract even virtually monolingual speakers across the language boundaries.

7.9 SUSY

SUSY (Saarbrücker Übersetzungssystem) has been in existence for a long time, and there is now quite a literature on it, but outside the University at Saarbrücken where it was developed it has not yet had any major application, though its influence has been felt, for example in German work on the Eurotra project (see below).

The present ongoing SUSY project dates back to 1972, and was at first concerned with translation into German from Russian. During the following years French into German and English into German components were added. Some work has also been done on Esperanto into German, Danish into German, Dutch into German, German into English and German into French.

Although work is going on with a view to using SUSY in various applications, such as SATAN (Saarbrücker Automatische Text-Analyse) for automatic indexing, SUSY-BSA (restricted text applications at the Bundessprachenamt), SUSANNAH (incorporation of outside terminology information) and MARIS (Multilinguale Aspekte von Referenz-Informations-Systeme), essentially it seems SUSY is still regarded as a research system.

The involvement of staff at Saarbrücken University in the SUSY project must also help them in the teaching of students and in preparing them for a world where machine translation will be the dominant form of translation activity. The importance of such an application of machine translation should not be under-estimated.

7.10 SOCATRA

This Montreal system produced by the company of the same name (acronym for Société Canadienne de Traduction Assistée) was created by the company's founder Claude Richard and operates in the language pair English to French (always the pair most in demand in Canada). According to the company's 1988 brochure translation speeds of up to 60,000 words an hour can be obtained with the system. The marketing of the system emphasises the use of Socatra's own system rather than the sale and purchase of individual system. Customers are invited to provide text for input to Socatra's main system; they can then receive either raw output, for post-editing themselves, or output which has already been post-edited by Socatra's own team of translators.

7.11 METAL

After being "imminent" for several years (a launch was promised at the Hanover fair in 1985, but there was then a pause), METAL is now available as an operating system, and one that may well make a major impact, firstly because it seems to have a large element of design modularity, so that it can be used for different language pairs; secondly, and unusually, because it has a European sponsor, and thirdly, because it is a system available now which includes some recent technology, in particular what is called fail-safe heuristics which will give an enhanced capacity for disambiguation. Throughput is more than 200 pages a day.

METAL (Mechanical Translation and Analysis of Languages) has in fact almost as long a history as any other system, the original project having been established at the Linguistics Research centre of the University of Texas, with United States government funding, as early as 1961, for the translation of German into English. The project was later terminated, after Government funding was reduced, but was again revived, and received a considerable boost when Siemens AG, of Munich, began to support the project. Since 1980 Siemens has been the sole sponsor, and developments in the system have been tested in Munich.

The oldest language pair in the system is German to English, and this is now commercially available; it was announced at the MT Summit Conference in Munich in August 1989 that there were now more than 12 installations. It is said to be particularly strong in the fields of data processing and telecommunications. Other versions being developed are French and Dutch language pairs at the University of Leuven in Belgium, English to German, originally developed in Texas, and German to Spanish being developed in Barcelona.

7.12 PC-Translator

One of the latest firms to enter the machine translation market is Linguistic Products, an American firm with a European sales base in Nice, France. They have caused some sensation because they are offering what they call a machine translation system, PC-Translator at $985 for the first language pair, a fraction of the cost of other systems on the market. They were also the first firm to bring out a product for translation between English and Danish, and a Swedish system is said to be in preparation. They have sold a number of their products for translation from English to French and English to Spanish.

As might be expected for the price, the system is linguistically unsophisticated, and only a few obvious pattern algorithms have been entered for each language. It is however designed to be user-friendly, and it is very easy to enter new terms and even whole phrases. It is obviously designed for use by the individual translator, but the crudeness of its target text output would slow such a translator down unless he was working with texts which involved considerable repetition.

However, many translators do work with such text, and with the system being offered at such a revolutionary low price it might prove an introductory system. Just as translators who have shied away from any automation in the past have found ALPS's Autoterm a most practical aid, and one that speeds up the translation of long documents to a considerable extent, perhaps to a factor of three, so too will PC-Translator help to revolutionise attitudes.

7.13 Globalink

This is another low-price system (around $2,000 for a language pair), but is more sophisticated than PC-Translator, and fits more in the mainstream of machine translation development. Language pairs offered are French, Spanish and German to English, English to French and English to Spanish. A few systems are said to have been sold in the United States (no figures have been published) but at the time of writing (September 1989) there had been no major marketing in Europe. A review of the system by Claude Bedard appeared in the September issue of *LT Electric Word,* but this appears to be the only published information so far apart from the manufacturer's brochure.

7.14 Gigatext

This company appears to represent an example of the problems which can arise in setting up a machine translation system from scratch, outside the mainstream of machine translation experience.

The Saskatchewan provincial government in Canada was faced in 1987 with a sudden huge translation requirement, as a result of a ruling by the Supreme Court of Canada in the so-called Mercure case, that, because it was a successor organisation to the North-west Territories, where in the 19th century both French and English were official languages, all the statutes passed by the provincial legislature since it came into being should have appeared in a French version as well as the English one.

The provincial government therefore welcomed an approach by what one newspaper described as "a flamboyant high-tech entrepreneur from Quebec and a maverick Manitoba computer professor with an unorthodox method of translating with computers".

The Saskatchewan government apparently put $4 million dollars into the scheme for automatic translation of the statutes, through a new company called Gigatext, but became concerned when no translations appeared long after the initial deadline. Expert advice has now been taken, and an investigation has been launched, as there may also have been some financial irregularities.

There is every reason to believe that the computer expert was quite genuine in his belief that he had a machine translation solution but outside experts who have been consulted do not think his system can work. The Saskatchewan government also seems to have been unaware, until too late, that there existed a mainstream of machine translation activity.

The lesson to be derived from this episode seems to be that there is an urgent need to create more publicity worldwide, not so much for machine translation as such, since that only encourages over-enthusiastic "experts" to follow their own inclinations, ignoring the lessons of 40 years of machine translation experience, but to publicise the fact that there is a mainstream of machine translation activity.

7.15 Tovna

The Tovna translation system was launched at the Translating and the Computer conference in London in November 1988, having been developed by a software expert, Dr Daniel Cohen, in Israel. Dr Cohen worked on developing his project as a research scheme throughout the 1970s, and did not obtain commercial backing until 1985. The first language pair to be developed is English to French, and a pilot system has been installed at the World Bank in Washington, USA. An English to Russian language pair is being developed, and there is interest from eastern Europe. A French to English system may become available towards the end of 1989.

Tovna claims to be an innovatory system in that it learns from actual use, i.e. it remembers modifications to its original suggested translation made by the operator and implements them for future texts.

Tovna was demonstrated at the Machine Translation Summit II in Munich in August 1989, when it also announced that it had secured an order from Lexi-tech Inc., of Moncton in Canada, where it would replace a Logos system.

Lexi-tech had been a new company established specifically to handle a $21 million Canadian defence sub-contract to translate and publish 100,000 pages of technical manuals for frigates (possibly the biggest single translation order yet placed). Lexi-Tech quickly acquired a dictionary of some 75,000 terms and expressions and, with its own software able to interface between machine translation and a desktop publishing program called Interleaf, it hopes to make Moncton an international centre for technical translation and re-publication.

7.16 Winger

Winger Holdings A/S, in Virum, Denmark, is a textile company which, in the process of computerising its own operations, became involved in software development. In particular it developed a database with a highly efficient method of storage and access. Their database is called Reduced Information Datastructure. The application for language translation seemed a useful development, and they have now produced (the launch is imminent) an English to Danish translation program, *Winger 92.*

So far little grammar seems to have been input into the system, although verbs are inflected, but homographs are recognised, and the operator is given a choice of possible meanings, and invited simply to delete those which are inapplicable.

The strength of the system undoubtedly lies in its database, but it will not become clear until there is some experience of practical application of the system how effective it is in speeding up translation.

7.17 Systems in Japan

Systems in Japan are, as might be expected, overwhelmingly for the translation of English into Japanese and Japanese into English, though a widening of the horizons is now occurring, and in particular there is great interest in translating between Japanese or English and the other major languages of South-east Asia, such as Chinese, Korean and Thai. There is a growing literature describing the development and composition of Japanese systems, but a dearth of information on user experience. *ATArishi,* the journal for translators between Japanese and English, published in the United States, carried a report of an independent translator's testing of the Fujitsu ATLAS system at the Fujitsu premises in Numazu, in its winter 1989 issue; the report details the progress of translation of a particular segment of text. The visitor paid tribute to the openness of Fujitsu but was unimpressed by the results he found.

According to some reports what was almost machine translation fever broke out in the first half of 1984 when Bravice introduced its Mini-Pack Japanese to English system to run under Unix on a DEC minicomputer, following it up a few months later with a microcomputer version, the Micro-Pack, at the then astonishingly low price of 0.65 million yen (approximately 4,000 ECU). Other companies then made haste to follow suit.

Meanwhile work was already well advanced on the Japanese government supported research project, the Mu Project, with which Professor Nagao was closely involved. This was a four-year project, 1982-1986, and has been succeeded by the Japan Information Centre for Science and Technology (JICST) project. Test operation of this system is scheduled to start within the next few months, and it may be possible to commence practical operation in 1900.

Commercial companies said to be marketing, producing or developing systems include (system names in brackets) Oki Electric (PENSEE), Canon, Sanyo Electric (SWP-7800), CSK, Sharp (DUET), Toshiba (AS TRANSAC - operational; TAURAS - experimental), Japan IBM, Systran Japan, NEC (PIVOT), Hitachi (HICATS), Fujitsu (ATLAS II), Bravice International (MicroPack), Matsushita Electric, Mitsubishi Electric (MELTRAN) and Ricoh (RMT).

It will be noticed that most of these companies are also computer manufacturers, and in many cases the machine translation software is specifically designed to run on that company's equipment. In some cases equipment and software is offered as a "work station".

Many of the systems were shown at the Machine Translation Summit II in Munich in August 1989, sometimes for the first time in the West. Most of the systems shown were for Japanese to English (Toshiba, Hitachi, NEC, Oki), or for English to Japanese (Sharp, NEC, Oki), but Fujitsu also showed a Japanese to German system.

If Bravice is disregarded, the market leader in terms of sales and sophistication of output is probably the Fujitsu ATLAS I (English to Japanese) and ATLAS II (Japanese to English) systems, which have been on the market for four years. Sales are now over a hundred.

Also shown was a prototype of one of the most ambitious Japanese products, the machine translation system for Asian languages, produced by the Centre for International Cooperation for Computerisation (CICC), which uses an interlingual approach (comparisons with Eurotra will be interesting when more information becomes available) for translation between Japanese, Chinese, Malay and Thai. The system is still under development, and only the Japanese dictionary (with 200,000 words) is near completion. The Chinese dictionary still only has 10,000 words, Malay 10,000 and Thai 5,000, but the developers are confident of reaching a translation speed of 5,000 words and hour.

There is also the ambitious collaborative project to create a Japanese Electronic Dictionary for English and Japanese which is being funded by eight commercial companies (including many of those named above) and the government, and is intended to save the participating companies much individual dictionary-building labour. The Japanese ability to co-operate when this is of mutual benefit gives them a further advantage.

Another ambitious research project is to develop a system for the automatic interpreting of telephone calls between English and Japanese, using speech recognition and speech synthesis. This system, started in 1984 and scheduled to take 15 years, will represent one of the most extraordinary advances if it can be successfully achieved, but little information has become available since the start of the project. The paper by Akira Kurematsu at the Munich Machine Translation Summit gave little indication of what progress had been made.

What seems to have happened in Japan, and this has world-wide implications, is that because of the relative inaccessibility to foreigners of the Japanese language, and because of what can only be described as their commercial dynamism, the Japanese are convinced that the way to future success in all domains lies through machine translation. The efforts they are making at both government and private level are massive. There seem to be none of the genteel reservations about machine translation which we nave in the West As Sigurdson and Greatrex, who estimate that some 800 to 900 people are currently engaged in research and development of MT systems in Japan (probably about 60% of these are working In commercial companies), put it: "No major technological breakthroughs have been achieved, but it would appear likely that the massive efforts in Japan would, in the typical Japanese way, produce a number of incremental results which would yield MT systems to be efficiently applied to a number of tasks."

All this must cause considerable concern in Europe, and in the United States. Because of the scale of the Japanese effort, their lead could become impregnable, and the post-industrial society, the Age of Information, into which we are moving could be dominated by Japan for decades.

There is concern that Japanese researchers are now working on systems which can handle, in addition to Japanese and English, French, German and Spanish, and that in a few years' time they may be able to market systems in the West.

However, there are indications that the Japanese to English systems which have been produced so far rely heavily on pre-editing, and are only suitable for use by Japanese users who can do the editing. This is because the Japanese language is a difficult one to analyse semantically, and sentences in natural written text are often highly complex.

Another indication is that many systems which have been bought by companies in Japan are lying idle, and this seemed to be confirmed by a survey conducted by Professor Makato Nagao, of Kyoto University, and reported to the Machine Translation Summit II at Munich.

Jon Sigurdson and Roger Greatrex. Machine Translation of On-line Searches in Japanese Databases, published by Lund University, Sweden, 1987.]

SECTION 8 RESEARCH SYSTEMS

There is probably no authority anywhere which is aware of all the machine translation projects going on in the world, and there may be interesting machine translation developments, such as Tovna and Winger described above, which will suddenly emerge from unsuspected sources.

The most comprehensive survey of research projects currently in progress was given by W. John Hutchins in his paper given at Budapest in August 1988.

Mention is made below only of those projects which are already known to be in development, and which are likely to prove of some significance.

8.1 GETA

France's main effort in machine translation has long been concentrated at the University of Grenoble and its CETA system, and later GETA (Groupe d'Etudes pour la Traduction Automatique) project. In 1971 the research team at Grenoble took the opportunity of a renewal of computer equipment to move to a transfer approach, which eventually emerged in the Ariane system, mainly used for Russian to French. Large-scale work was also done on German to French. The work done on Ariane and its operating model offshoot PROTRA enabled the researchers to work with various other language pairs with researchers in various countries not previously associated with machine translation systems, including Brazil (Portuguese), Thailand (Thai) and Malaysia (Malay).

GETA has also been closely involved in the French national project for machine translation, which developed into the Calliope project. A documentation company, Sonovision, later SITE, which specialises in aeronautics manuals, cooperated in the development of Calliope-Aero for aeronautics French to English translation, and this in turn led to Calliope-Info for English to French translation of data processing material. There was also a project to develop a translator's work station, Calliope-Révision.

8.2 DLT

The DLT (Distributed Language Translation) system is being developed in the Netherlands as an interactive multilingual system for text generation over computer networks, with each terminal acting as a translating machine from and into one language only, but with an intermediary language, based on Esperanto, being used for inter-terminal transmission. It is therefore intended more for multilingual communication than for machine translation, although it uses machine translation technology. The makers, BSO of Utrecht, demonstrated a prototype model in December 1987.

8.3 Rosetta

This is an experimental project being carried out by Philips of Eindhoven in the Netherlands, and thus is one of the few examples outside Japan of a major industrial company becoming involved in machine translation (other examples are Siemens and IBM). The project began with experiments in 1980 and until 1985 was only a small research activity. In the latter year, however, it was expanded and experimental translation was set in being between Dutch and English and Dutch and Spanish. This phase, Rosetta3, is now drawing to a dose and the researchers are expected to start work soon on Rosetta4, which is designed to lead to an operational system.

8.4 JFT-IV

The People's Republic of China has for the last 13 years been developing the JFT system, designed for translation into Chinese from English, French, German and Russian, and at present is concentrating on English to Chinese. The current version of the system is JFY-IV. An experimental model is also being established to translating from Esperanto to Chinese, Esperanto being a language highly popular in China.

8.5 LMT

In the early days of machine translation development a major role was played by IBM, and in view of the involvement of so many Japanese computer and electronic equipment manufacturers with machine translation it might have been expected that IBM would be a major player. However, it does appear that IBM are content to maintain a fairly low-key research programme. Their main research system is LMT (acronym for Logic-programming-based Machine Translation), and the system was described at the Machine Translation Summit II in Munich in August 1989. Prototype versions have apparently been developed for English to German, English to Danish, Danish to English, English to French, German to English and English to Spanish.

IBM of course does have a huge translation requirement, and as far as European languages is concerned, while the actual translation is usually performed in one of the countries of the target language, the whole effort is co-ordinated by IBM European Language Services division in Birkerod in Denmark. This division is about to introduce a Translation Support Facility which it has designed itself, which is a multi-tasking work station allowing the translator not only to call up dictionaries and glossaries on-screen (thereby imposing IBM preferred vocabulary where different translators are working on the same text), but also to identify and retrieve previous translations where there is repetition of a segment of text.

8.6 Eurotra

The Eurotra project of the European Communities is currently the largest single machine translation project in the world in terms of number of personnel, and possibly in terms of expenditure. The aim is to lay the foundations for the development of a system for multilingual high-quality translation between all the official languages, Danish, Dutch, English, French, German, Greek, Italian, Portuguese and Spanish.

For political reasons research has been decentralised to units in all the countries of the Community, with only a comparatively small central unit, and this decentralisation has made administrative co-ordination difficult The project has not been without its critics, and it is now obvious that it is still a long way from producing an operational system. Nevertheless, the much criticised decentralisation has had considerable success in promoting computational linguistics research in Europe, especially in the smaller countries. In Portugal, to give one instance, a new *Instituto de Linguistica Teórica e Computational* has been established as a direct result of the involvement with Eurotra, and this institute is developing other projects, and is organising in December this year Portugal's first conference on computational linguistics.

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 co-ordinated 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. 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.

8.7 Other countries

There has been little recent information about developments in the Soviet Union and other East European countries, although further information may be forthcoming at the "Computers and Translation 89" conference to be held at Tbilisi, Georgia, at the end of November 1989.

Saudi Arabia appointed two part-time machine translation researchers to assist on an machine translation project in 1983, and a term bank has been established. But here and elsewhere in the Arab world interest in machine translation is mainly confined to preliminary research.

SECTION 9 MARKET SHARE

According to the Bureau Marcel van Dijk's report, the proportion of the European market held by the various suppliers is: Logos 35%,, Weidner 25%, ALPS 15%, Systran 10% and others 15%, but the source for this calculation is not stated.

I agree entirely with one of the comments made in that report: "Le choix de I'un ou de I'autre systems est souvent le fruit du hasard ou d'une opportunité, rarement le resultat d'une étude comparative des performances. En effet, peu d'utilisateurs connaissent le marché de la traductlon automatique, et notamment les produits concurrents de celui qu'ils utilisent."

The reason for this apparent naivety, even on the part of multinationals, almost certainly lies in the ignorance of the problems inherent in natural language processing, and the automatic assumption that "a computer can do it". Every effort must be made to educate potential customers for machine translation on the complexity of the task, and the need for thinking it through.

Bureau Marcel van Dijk, Etude des Strategies de Financement et de Commercialisation de la Traduction Automatique, report for the Commission of the European Communities, June 1988.

SECTION 10 DOCUMENTARY MATERIAL

There is already a considerable volume of literature on machine translation, and this is being added to daily. The purpose of this section is to indicate where and how experiences in the applications of machine translation are being recorded, and to make recommendations for further co-ordination.

10.1 Books and bibliographies

Extensive bibliographies of machine translation can be found In the Hutchins 1986 history (supplemented by the bibliography following his 1988 Budapest paper), and in Slocum. The Canadian Translators and Interpreters Council have an astonishingly comprehensive database on everything written on translation, including machine translation, provided it involves Canada.

Mr W. John Hutchins, who has already been mentioned on several occasions, has for some years now attempted to assemble all published matter and unpublished documents relating to machine translation developments. Although Mr Hutchins is deputy librarian at the University of East Anglia, Norwich, this work is being done not through the university but on his own initiative, although of course he is using his librarianship expertise to identify, locate, assemble and index material. The archive, which has been computer-indexed, and which takes up a whole room at his home, is thought to be the most comprehensive anywhere in the world. However, with the increasing pace of machine translation development, a single individual may have difficulty keeping up with all new material, and it is recommended that efforts be made to ensure worldwide co-ordination in spotting and collecting new material.

W.J. Hutchins, Machine Translation, past, present, future, published 1986 by Ellis Horwood Limited, Chichester, UK.

W. John Hutchins, Recent Developments in Machine Translation, published in Dan Maxwell, Klaus Schubert and Toon Wttkam (ed.), New Directions in Machine Translation, published 1988 by Foris, Dordrecht, Netherlands.

Jonathan Slocum (editor), Machine Translation Systems (Studies in natural language processing), published by the Cambridge University Press 1988 (previously published as special papers on machine translation by the Association for Computational Linguistics, 1985).

10.2 Conferences

Conference proceedings are a major source of information, perhaps the most important source, and the conferences themselves provide unique opportunities for those active in the field of machine translation to meet.

In 1977 the Commission of the European Communities organised a conference in Luxembourg on the subject *Overcoming the language barrier.* This conference acted as something of a catalyst in bringing together for the first time machine translation developers and researchers, users, and translation practitioners. Although the conference was also ostensibly the third European congress on information systems and networks, it was undeniably innovative, and its proceedings are still a major reference source.

The following year, as a direct result of this initiative, the UK information science organisation Aslib, in conjunction with the Translators' Guild of the Institute of Linguists, held a conference on Translating and the Computer", which became established as an annual series (the Institute of Translation and Interpreting has taken over from the now defunct Translators' Guild in co-sponsoring the conference).

Another important series, though concentrating very much on the theoretical side of the question, have been the COLING (International conference on computational linguistics) conferences, held in Ottawa in 1976, Bergen in 1978, Tokyo in 1980, Prague in 1982, Stanford California in 1984, Bonn in 1986, and Budapest in 1988. The next conference in this series is to be held in Helsinki in 1990.

The series of conferences held at the Cranfield Institute of Technology on the methodology and techniques of machine translation, Cranfield Institute of Technology seem to have come to an end.

The World Systran conference in Luxembourg in February 1986 was a major event in the development of machine translation, and provided a unique opportunity for the exchange of users' views.

The first so-called Machine Translation Summit was held in Tokyo in September 1987, and the second in Munich in August 1989. The third is to take place in North America in 1991.

In April 1989 the Japan Electronic Industry Development Association acted as host to an International Forum on Translation Technology (IFTT 89), on the theme of harmonising human beings and computers in translation.

Machine translation is of course often a subject at the conferences of translator organisations, particularly in Canada (CTIC), the United States (ATA) and the United Kingdom (ITI), less so in other countries. A number of major papers were given at France's largest conference so far on translation topics, the *Journées européennes de la traduction professionnelle,* held in 1987. Machine translation as a subject Is also dealt with at the seminars at the language shows which are now burgeoning in Europe, and in 1989 has appeared specifically on the agenda of Expolangues, the London Language Show, and Expolingua Portugal.

Machine translation is sometimes dealt with as a subject at information industry conferences, such as EURIM, and the natural language processing of speech was a subject at the first Eurospeech conference held in Edinburgh in 1988, and at the second conference held in Paris in September 1989.

10.3 Journals

In the early years important works on machine translation appeared in such publications as *Computers and the Humanities,* and this prompted the launch of the theoretical journal *Computers and Translation.*

The British Computer Society Natural Language Translation Specialist Group Newsletter and the ALLC Bulletin have also carried articles which have had some impact

Language Technology, recently renamed *LT Electric Word,* has a wide-ranging brief which covers almost any form of text processing, but covers the machine translation scene from a news-reporting point of view, though occasionally it has in-depth articles, such as one on speech recognition.

Language Monthly (1983-1988) and *Language International* cover language activities in general, but this has included information, again from a news-reporting point of view, on machine translation.

Other general language or translation magazines, such as *Meta, Jerome Quarterly* or *Lebende Sprachen,* and the journals of the translator associations have carried occasional articles on machine translation; some of the early work on TAUM, for example, was written up in *Meta.*

10.4 Theses and dissertations

Although Mr Hutchins attempts to collect theses and dissertations submitted for higher university qualifications, it is often difficult to identify these, and it would be useful for some form of co-ordination to be established. There are organisations in the United States and in the United Kingdom, and possibly in other countries, which will identify and obtain copies of theses and dissertations on particular subjects.

10.5 Patents

Similar remarks apply to patent applications as to theses and dissertations.

10.6 Surveys

The Commission of the European Communities has commissioned from Ink International and *Language Technology,* of Amsterdam, a survey of the language industries of Europe, and this will serve as a basis of monitoring developments, particularly in the commercial sector. At the same time a similar survey is being carried out in the French-speaking countries *(Francophonie)* of the world under the aegis of the *Association pour Cooperation Culturelle et Technique* in Paris. SCS Informationstechnik GmbH is currently carrying out a survey for the German Federal Ministry for Research and Technology *(Bundesministerium für Forschung und Technologie)* on the state of the art of, attitudes to and prospects for machine translation.

10.7 Associations

At the recent conference in Tokyo, IFTT 89, the suggestion was made that the time had come to form an international association for machine translation; such an association could contribute much to coordination in this field; the suggestion was echoed by several speakers at the Machine Translation Summit II in Munich in August 1989.

Another function of such an association could be to act as both a public relations body and as a reference point on machine translation, since naive assumptions are often still made, even by large corporations and public bodies.

SECTION 11 TEXT TYPOLOGY

11.1 Literary texts

For 2,000 years, from Cicero to the German *Übersetzungswissenschaft* school, those who thought or wrote about translation had largely in mind what may be broadly described as literary translation, that is the translation of rhetoric, poetry, drama and fiction.

It is a commonplace, in any writing about machine translation, to comment that it is of course unsuitable for literary translation work, before moving on to discuss more promising fields of application.

It might, however, be useful to consider the possible application of the computer in the field of literary translation, as the reasons for its unsuitability, if valid, might throw light on its suitability or unsuitability for other applications.

The reasons why the computer is thought to be unsuitable for the generation of literary translation is that the words used in literature can be rich in connotations, whereas the computer operates on the basis of straight equivalences; that literary translation requires an in-depth knowledge of the culture behind the source language text, as this may require interpretation and elucidation - the translation needs to be enhanced where necessary - whereas the computer has only the words of the source language text to work on; that the target language needs to be manipulated imaginatively rather than literally; and that in literary works style and register are part of the "message" to be conveyed, and such sophistication is beyond the computer.

It is possible to take a passage from Shakespeare, and show that the choice of words is informed by a multitude of stimuli and cross-references, and that the whole line of verse is cadenced to give a particular effect. Looking at the cultural angle, it can also be argued that no-one can really understand Shakespeare, or Milton, who does not take into account the world picture they had (to take a simple example, in our post-Copernican thinking, we find it difficult to appreciate the different way of thinking of people who thought that the Earth was the centre of the universe); or in Eugene Nida's celebrated example, the concept of the Lamb of God, with its implied innocence, means nothing to the Eskimos.

We ought, however, to take into account the work that has gone on in the application of the computer in literary criticism. That this is considerable is shown by the existence of an association and a journal (Association for Literary and Linguistic Computing, *Journal of Literary and Linguistic Computing),* as well as regular conferences on the subject. The British Centre for Literary Translation which has recently been founded at the University of East Anglia is considering undertaking research on lexicographic work for translators, and the idea of a multilingual thesaurus has already been mooted as a possible project

Computers can be used to analyse style, and find concordances. Concordance programs are widely used, for example, in Bible studies. Surely such programs could be converted into tools to assist the literary translator.

This brings us on to the point of dictionary look-up. It has long been realised that dictionary look-up is one of the most time-consuming of tasks for the human translator, and much thought has been given to the speeding up of this task.

However, the assumption is often made that translators, as a general rule, look up the words that they do not know. This is a fallacy, but a widely-believed one. As many translators themselves have not analysed how they work, there can be little surprise that the layman thinks that dictionary look-up is to find unknown words; so there is an assumption that the way to approach automation of this aspect is to locate "words not found" and to "paste" equivalences into the translation.

This bears out the point made several times at the November 1988 Translating and the Computer conference, in London, of the need for a model of how the human translator actually works.

In fact, experienced translators working in their own subject field do not come across all that many words they do not know, that is to say, of whose meaning in the source language they are ignorant. Of course much translation is done by persons who may be ignorant either in language knowledge or subject knowledge, and machine translation has to cater for this. We deal with this point later under typology of translation activity.

But as literary translation, as is technical translation done by a specialist translator, is usually performed by persons with adequate subject and language knowledge we can look more closely here at the phenomenon of dictionary look-up.

Dictionary look-up is more often than not used by the translator to jog his mind; he knows what the word means, but cannot immediately call to mind a suitable target language equivalent. He will come across "baanbrekend" in a Dutch text, but for some reason the word he wants, "pioneering" in this particular context, does not swim into his mind - but as soon as he sees it in the dictionary he knows that is the word he wants.

Translators, it has been found, look up more words when they are tired than when they are fresh, because when they are fresh their minds are more ready to flick through their own target language word store.

This aspect of dictionary look-up applies to all translators, but is especially important for literary translators. What they require, however, is not so much dictionaries of equivalences, such as are provided by many data bases available for use on screen, but something like a thesaurus (many literary translators make considerable use of books such as Roget's Thesaurus), where they can go on a word chase, tracing an idea or a concept through synonyms and near-synonyms. It would be a boon to literary translators if some sort of concept-based synonym-rich program of this sort could be offered, perhaps similar to the Ericsson technical terminology program, reported in *Language Monthly* in 1988. It would also be useful to writers.

Terminology databank based on conceptual relationships, article in Language Monthly, no. 55, April 1988.

11.2 Bible translation

A huge effort is still devoted every year to translating the Christian gospels, and it should not be forgotten that the original Inspiration of the Brigham Young University program in the 1950s was to provide means of translating Mormon gospels into other languages. Much of the Bible is actually narrative of a reasonably straightforward kind, and in fact there is no reason why it should not lend itself to machine translation. The problem is that the largest demand for Bible translation is into obscure languages, but it is not inconceivable that a future time will come when there will be some automatic or at least highly user-friendly system for analysing the structure of target languages so as to enter parameters for semantic analysis and language generation routines.

11.3 Legal translation

The translation of legal texts between languages sharing a common legal culture and having linguistic similarities, as is the case for French and Italian, or Spanish and Portuguese, should not present too many difficulties. But translation across cultures, such as between English and German, or between English and Chinese (a current major problem in Hong Kong), presents large problems even to the human translator. The problems result from the frequent lack of correspondence between legal systems, or national practices or ways of looking at things. A case often quoted is the difference between the English and German approaches to insolvency (intermediate forms of court jurisdiction can occur at an earlier stage in German practice, whereas the UK tends to place more onus on the debtor to sort out his own problems), or the difference between English rules on what sort of evidence is admissible in court - hearsay evidence, for example, is usually not allowed - and the Chinese habit of allowing full circumstantial explanation.

However, although in the past many writers on the subject have declared that machine translation is not appropriate for legal texts, there is a need for large-scale translation in this field. We have already noted the requirement for Saskatchewan province in Canada to translate nearly 100 years of legislation into French (the province of Manitoba has the same problem) and the requirement to translate much Hong Kong legislation based on English law into Chinese. In addition there is the ever increasing amount of laws and decrees being issued in the European Communities, and in Belgium there is an unsatisfied demand for translation between French and Dutch of legal commentaries and digests.

Also, particularly in commercial law, judgements (not to speak of evidence) can run into scores and even hundreds of pages.

There is also in legal translation a considerable proportion of repetition, and the use of stock phrases. Machine translation can also help to secure consistency.

The conclusion must be, therefore, that while legal translation is not an easy task for machine translation, it is one that ought to be tackled, since the very fabric of international relations may start to unravel if translation cannot keep pace with the intensification of translingual transactions.

11.4 Financial translation

There is a considerable amount of financial translation performed in Western countries- some freelances do nothing else. This requirement will certainly increase after the coming into being of the Single European Market in 1993. Much of the work is concerned with company reports and balance sheets, budgets and estimates. There is likely to be an increase in work of translation of summarised company information (for company searches etc.), where machine translation could usefully be associated with information services. Provided some way can be found of dealing with the awkward formats and plethora of figures (Systran is equipped to tackle balance sheets) this sort of work would seem to lend itself to machine translation, as much of it is repetitive, and most machine translation systems can cope with well-written abstracts. One disadvantage is that really long documents are rare, but since documents do tend to run to a pattern this need not be so big a disadvantage as it appears.

11.5 Scientific translation

The biggest requirement in this field is for the cover-to-cover translation of scientific journals, especially in physics, chemistry, the various branches of medicine and pharmaceutics, and defence procurement.

There is also a continuing need for the scanning of scientific journals to see if there are items worth earmarking for subsequent full translation, and of course machine translation has been involved in this field almost from the beginning.

The main problem is the fact that many of the articles concern leading-edge technology, with new terms and concepts often being used. If machine translation is to be applied extensively in the field of scientific translation, it may have to be customised to the needs of individual sectors, and even individual translators.

11.6 Technical translation

This is a field which is well documented, and can be the most easily adapted to machine translation. The complexity of modem products requires ever longer operating instructions, maintenance manuals and workshop manuals. Often the writing of such manuals is done on an ad hoc basis, but the professions of technical writer and illustrator are now becoming organised in many countries. Such persons are going to be of increasing importance to machine translation, as the feasibility of machine translation depends to a large extent on how the source text is written. This point is dealt with in greater detail in section 12.

Another possible application of machine translation in this field is specification scanning. Many firms look at a whole range of calls for tenders, and specifications, only to decide on making a bid in a minority of cases. Raw machine translation could be of utility here.

11.7 Patent translation

Patent translation is often regarded as a branch of technical translation, but in fact patents are more and more a matter for the specialist translator, since they are written in a particular manner, and written for particular objectives, to which one needs to become accustomed.

Patents are notoriously difficult to translate, since the patent agents have over the years become adept at using words, almost speciously, to gain as much ground as possible - widening the scope of the patent as much as possible but not so far that it will be thrown out by the authorities as being too widely drawn; and explaining sufficiently to show the originality of the invention while giving as little away as is consonant with a successful application.

Sometimes patents need to be translated for filing in another country, and such translations have to be done with great care. But sometimes patents are translated only for information purposes, in which case often only the claims, and not the description, are translated.

Patent documents, too, are not usually of great length, often comprising between eight and 20 pages.

11.9 Summaries and abstracts

There is a large latent demand for combined translation and summarisation, but automatic summarisation is still considered technically a highly difficult problem.

SECTION 12 INTEGRATION WITH DOCUMENT PROCESSING

The most successful commercial applications of machine translation have been where the technical publication departments of companies have been able to control processes upstream and downstream of the translation activity.

In the case of Rank Xerox and Perkins Engines this enables the company to impose a form of controlled language on its technical writers, which makes texts suitable for machine translation; according to papers given at the IFTT 89 conference in Tokyo in April 1989 this sort of approach is spreading in Japan.

In the case of Lexi-tech Inc. in Moncton, Canada, the machine translation system has been integrated into an automated document processing system, including the sophisticated handling of graphics. Captions which appear in English on drawings can be taken off the drawing, automatically translated into French, and the French text then placed back on to the drawing.

This has a number of implications, on the training of translators, for example, on the sort of translation service company we can expect to see emerging in the next decade, on project management, on the future of translation companies, or even on how translation is organised in international bodies such as the European Community. As more and more companies and organisations become involved in the production of multilingual documentation on a large scale, technical publications departments or document production departments are likely to assume greater and greater importance. Translation is likely to be one, but only one, of the services of such departments.

SECTION 13 SUITABILITY OF TEXTS FOR MACHINE TRANSLATION

How does one go about deciding which type of text is suitable for machine translation?

The time will come when machine translation will be of assistance in any translation, and will be a commonly accepted tool just as the word processor is now; that time is not so far off; but for the present it is obvious that some documents are more suited to machine translation than others.

One criterion is obviously length. Practical experience seems to suggest that it is not worth putting documents of less than 30,000 words through a system. This precise figure was quoted to the author of this report on successive days, by the team at Multiscript/La Langagerie in Montreal in Canada, who were using only the terminology support system on ALPS, and by Mr Larry D. Rogers, systems manager at Lexi-tech, Moncton. Individual documents may of course be of a shorter length than 30,000 words, provided that they are of a similar nature to what has gone before, using similar terminology, and the work flow is uninterrupted.

Other criteria are obviously subject matter, complexity of both language and layout, and the requirements for consistency and speed.

Control of as many processes upstream and downstream of translation as possible - writing of the original text, inputting of the original text, translation and post-editing, and outputting of the final text in its definitive form is also conducive to the cost-effectiveness of machine translation, which suggests a closer link-up with technical writing on the one hand, and electronic publishing on the other.

An important criterion is that of final application. One huge requirement, which is often ignored by the conventional translation world, but one where machine translation has demonstrated its utility, indeed its indispensability, is that of information scanning, as has already been mentioned at various points.

In the commercial, financial, technical and scientific fields of today no-one reads every word of every document which passes over their desks; quite the contrary: the trend is to scan more and more documents, like panning for gold, in order to identify really usable information.

Even translation companies are finding that more and more of their work is scanning rather than translation. A frequent requirement is looking through trade publications to find matter of relevance to a customer, in the light of the customer's briefing, checking for invitation to tender announcements, looking through patents being filed; or an initial examination of a specification to see if it is worth a company tendering.

The other aspect is the "read-through" (term used by the Science Information Reference Service at the British Library in London, which provides this service), where the translator sits with the customer and indicates, from titles and headlines, what a particular document is generally about.

The greater part of information from such sources will not be relevant, but there may be gold dust there somewhere for the customer.

How much more useful it would be if the whole document could be translated, however crudely. So it is not surprising that raw machine translation is used for this. The United States Air Force has used Systran for gaining information about Soviet defence technology since the early 1970s. The expert in the field can quickly work out the information being conveyed, even if the language is not as idiomatic as the conventional linguist would find acceptable, and what in fact happens is that the specialist begins to "understand", or become fluent in, the raw translation style.

At the London Language Show (June 1989) the proprietor of a modest-sized translation company, Comtex, which has been using Weidner on an experimental basis, told the author of this report that she had customers whose main requirement was for raw translation.

SECTION 14 TYPE DEVELOPMENT

It is in fact instructive to note that the systems most in use today, Systran, Weidner, ALPS and Logos are, in origin at least, based on the 1960s technology which had so failed to impress the authors of the ALPAC report Of course these systems have been improved since, largely by the writing of additional paradigms. But developments in the last three decades, with more detailed analysis of deep language structure, informed by advances in linguistics, mathematical knowledge, bigger memories and term data bases of much improved quality, and the application of artificial intelligence in such fields as homograph resolution and context definition, have not, or at least have not yet, substantially altered the way computer translation is applied. There has been considerable reliance on the ability of computers to compute at very high speeds, and therefore solving problems by persistence, exploring all possible solutions to find the optimum match.

New elements are now being introduced with some of the systems now coming on to the market, such as METAL, Tovna, and Japanese systems. Even a comparatively small project such as Winger may make a contribution through its efficient database handling.

The Eurotra project and other projects in development such as DLT may lead to a movement to interlingual systems; on the other hand the increasing modularity of older systems such as Systran may show up the interlingual approach as being too cumbersome; it is too early to say.

The application of voice recognition and voice synthesis, which could eventually have a major importance in machine translation, is still at the research stage. So too is teaching computers to learn. Learning can consist of automatic analysis of feedback from users so as to correct errors, and analysis of text corpora in order to make correct inferences.

There is as yet no single obvious route for future development, and much to gain from retaining a multiplicity of systems and system developers. Seen from this point of view it may not even be wise to put too much pressure on Eurotra to evolve an operational model, provided that experts can be satisfied that relatively unfocussed research work is still capable of yielding results.

SECTION 15 THE TYPOLOGY OF TRANSLATION ACTIVITY

In this section we examine where translation activity is carried on, and to what extent it is currently automated.

15.1 In international or government organisations

The two major centres of translation activity on an international scale are the European Communities and the United Nations Organisation. The European Communities has a thousand or so translation staff, mostly located in Brussels and Luxembourg, now with their own Directorate-General. In addition machine translation work is coordinated through Directorate-General XIII.

According to a report of a "Carrefour de réflexion" on machine translation held at the European Communities in September and October 1988, machine translation accounted for only 0.5% of the total workload of the translation service. One reason for this low figure was obstacles in the way of providing the source text in machine readable form, and another the difficulty of interfacing with the word processing computers in the receiving departments.

According to a paper given at a conference in New York in 1985, the translation division of the United Nations in New York handles approximately 180 million words of translation every year (figures up to 1984, expected to rise 20 per cent in 1985). The total headquarters translation staff is approximately 400. Although the paper referred to was entitled "Computer-assisted translation at the United Nations", the extent of the computer-assisted translation seems to be in the provision of word processors to the more fortunate among the translators, who are also being given access to terminology databases.

The largest single national government translation body in the world is almost certainly that of the Translation Bureau of the Canadian Secretary of State, with, by some counts, some 1,800 translation staff. The bureau is seriously concerned that in the years to come it will be unable to cope with the demands made on it unless it automates the translation process. A recent government report, the Charpentier report, published in March 1989, forecasts a shortfall in the number of trained university translation graduates in the next decade.

Few other government bodies in the world seem to have large bodies of translators available in one place. The Japanese government, as we have seen, gives considerable encouragement to machine translation projects, but we hear little about government translators themselves. One of the liveliest organisations in Europe is the Bundessprachenamt in the Federal Republic of Germany, but no recent details appear to have been published of how many translators it employs or the degree of automation.

The most detailed across-the-board study anywhere of translation activity in government departments is almost certainly a 1985 monograph by Ted Crump, which examines each United States government department in turn. The only references to machine translation I could trace were for NASA and the Wright Patterson Air Force base. However, Harry Obst, director of language services for the Secretary of State, is currently looking at machine translation systems, according to a personal communication to the author of this report at the London Language Show in June 1989.

Report of the Carrefour de réflexion on Machine Translation, 26 October 1988, Commission of the European Community.

Nigel Cassar, Computer-assisted translation at the United Nations, paper given at the fourth annual conference of the Center for Research and Documentation on World Language Problems, New York, December 1985 (proceedings published by the Center, 1986).

Ted Crump, Translations in the Federal Government, 1985. published by Ted Crump, Washington, 1985.

15.2 In commercial and industrial companies

There is no pattern to this form of activity. Some companies, though relatively few in number, have properly structured translation departments (Shell, Krupp, Ruhrgas). In many, translation is an adjunct to either the technical publications department (Perkins Engines), or to the sales department. But the prevailing pattern, which one feels cannot much longer survive, is to solve problems on an ad hoc basis, putting work out to translation companies, recruiting freelances directly, finding someone in the company to tackle the work, entrusting the work to agents in the country of the target language. The Internationale Vereinigung Sprache und Industrie has chosen as the subject for its 15th annual conference, to be held in Nottingham, England, in November 1989, "Corporate language policy", in an attempt to persuade business that every company should think now about how it will handle language and documentation (from technical writing of the original documents, to electronic publishing) now that business is becoming multinational and multilingual. The first London Language Show, held in June 1989, also targeted the business user of language services, and the 1990 show will have a companion show at the same time, the Euro Communications Show, to include documentation processing more generally.

The degree of automation in companies is also patchy; some have installed machine translation and the more successful have recounted their experiences at conferences or in journals. But many others, even when they have a steady throughput of work (such as automobile manufacturing companies), are still working from hand to mouth. It is suggested that there is a great deal of promotion work to be done, but that it would be counter-productive to concentrate solely on the language side of machine translation; machine translation should be marketed as part of a documentation processing re-organisation.

15.3 In translation companies

There has been a tendency for translation companies to become larger, either by growth or amalgamations. In the last five years we have also seen, for the first time, interest from larger companies in other fields - Mather and Ogilvie, the advertising agency, acquired Euramerica. the largest company in the United States, while in Europe the Dutch publishers Kluwer acquired Ink International. ALPS, which for many years was primarily a machine translation and software products company, as already reported, have acquired Multiscript-La Langagerie, Interlingua-TTI and Haehl. Multiscript-La Langagerie, in Montreal, Canada's largest company, was already the result of an amalgamation, as its name implies. So too was Interlingua-TTI in the UK, to be known, from May 1989, as Interlingua. One of Europe's largest translation companies, Randall-Woolcott Services, of Gerrards Cross, England, has acquired recently a number of specialist smaller companies. What was once a small UK provincial translation company, Liverpool Translations, has embarked on a buying-up spree, and has already acquired Spargnapani in Manchester, one of England's oldest translation companies, and Ken Pull Associates in Essex; Liverpool Translations are likely to change their name to something more international-sounding in the coming months. A consortium of translation companies from every EC country has recently been formed to carry out work placed out by the translation departments of the European Communities. A group of small companies in northern Italy recently came together for a joint marketing exercise at the Expolingue Milan language exhibition. The language school company Berlitz has bought up this year (1989) two of the largest translation companies in the United States, and one of the largest in Scandinavia, Institut for Fagsprog in Copenhagen. Ink, Netherlands, is carrying out a deliberate policy of Europe-wide expansion.

The trend therefore is clear, towards bigger operating units.

Such a trend will obviously facilitate greater automation of the translation process. In the past many translation bureaux operated by putting the bulk of the work out to freelances - acting in fact as agencies. The complexity of translation activity is making this an increasingly unworkable solution, and it is noticeable that two companies which have seen perhaps the greatest expansion in recent years, Randall-Woolcott Services (UK) and Ink International (Netherlands), do practically all their work in-house. Such a trend will obviously favour the introduction of automation.

One of the obstacles to implementing machine translation in translation companies in the past has been that work received is so diverse, in format, application and subject matter, as well as in language combinations; but ail translation companies, large and small, confirm the same trend: jobs are getting bigger, and jobs usually involve some other process or processes in addition to translation. Source texts are also, at last, beginning to be received in machine-readable form.

It should also be mentioned that in the past freelance translation rates paid by translation companies, especially for the common language combinations, have been, in most countries (Sweden is a notable exception), very low. This has meant that machine translation, with all the investment involved, never seemed to be cost-effective. Good freelance translators are now finding they can charge a premium, since almost without exception they are overloaded with work, and so the machine is beginning to look more attractive.

So far there has been little use of machine translation in translation companies - a notable exception is TAO International in France. Even the well-publicised experiment by Mendez, of Belgium, of editing in-house raw Systran output received by telecommunication, seems to have been discontinued. The ALPS-acquired companies have all apparently successfully implemented the Translation Support System, and once-sceptical translators have been converted. ESC in the UK were at one time reported to be successfully using Weidner in a translation bureau situation, but this company may now be having financial problems. Lexi-tech Inc., in Canada, already discussed in some detail, cannot perhaps be regarded as a typical translation bureau, unless we are to accept that in future there will be a differentiation between the large organisation working on major projects only, and the translation bureau of the conventional type reacting to changing customer demand.

The pattern, then, with translation companies is patchy, but there does seem to be some movement towards machine translation.

15.4 Individual translators

There is considerable variation among freelance translators. There are some who have a bigger turnover than some translation companies, while others are mere part-time dilettantes. But many freelances are definitely interested in machine aids to translation, and in translation itself, as can be seen from the attendances at conferences in England and France where the subject is dealt with. The dominant role played in translator organisations in some other countries by persons who show little interest in the new technology is perhaps unfortunate (it explains why one-third of the attendance at the UK Institute of Translation conferences is from the other EC countries), and is certainly building up considerable frustration, but it cannot last.

So far there have only been tentative experiments by translators even with machine aids such as Ink Text Tools or Termex, but every seminar which the Institute of Translation holds on these is always a sell-out. I foresee considerable interest too in the PC-Translator being offered by Linguistic Products. Another possibility is that soon translators may begin to take raw translation on-line for post-editing in-house.

At the moment the biggest problem for the freelance in automating his translation process is the inputting of the source language text, and it is a matter for surprise that at a time when there is a photocopying bureau in every village street, there are no Optical Character Reader bureaux. This is almost certainly something that will be rectified in the next few years.

15.5 Para-translators and Non-translators

Until recently the greater part of translation work in the private sector was not done by persons for whom translation was their full-time career, but by persons who did translation only occasionally, as an add-on to their normal jobs. The proportion is difficult to quantify, but is still large. Many would-be users of translation services are still not aware of the mainstream of translation activity, or how to get in touch with it. Those who are engaged full-time in language activity may fancy that we have a high profile, but in fact language does not occupy the attention of the world as it ought to.

Commercial companies, previously unacquainted with the language world, who hear of machine translation, immediately assume that this means fully automatic translation, and are often deterred when they learn that they need to employ linguists. Perkins Engines, of Peterborough, by careful organisation, has managed the unique achievement of successfully using machine translation without employing linguists (except in the initial stages as consultants). The DLT system in Utrecht is being produced with non-linguists in mind.

The lesson is, that if translators as an organised profession turn their back on the new technology (rather than be constructively critical), then the world will still use the new technology, but without the translation profession. This would be a pity, since translingual ability and language sensitivity has a great deal to contribute even when the translation process has been largely automated. There is still a considerable promotional effort to be made here, particularly with some of the less forward-looking translator associations, and with the Fédération Internationale des Traducteurs. At the end of October 1989 the Italian translators' association, in a forward-looking move, are organising a conference on the new tasks (such as post-editing machine translation) opening up to translators.

If the future of Europe, and indeed of the world, is to be multilingual, then there must be an exponential increase in the availability of translation. Translation must be ubiquitous. Every businessman must know that he can get translation of his material done quickly, and conveniently. Translator organisations must get rid, and get rid quickly, of the notion that they are a sort of medieval guild, preserving a craft which should not be practised by impious hands. The customer must not be faced with a negative "can't do" attitude; "no problem" must be the response. No-one should be encouraged to think that translation can be obtained by pressing a button, but at least Germans and Italians, for example, should be given the experience that it is much easier for machines to turn their German or Italian documentation into English than it is for themselves and their staff to struggle with English.

SECTION 16 POSSIBLE AREAS OF DEVELOPMENT

The machine translation world at the moment is being pulled two ways - the temptation to concentrate on developing comprehensive flexible, intelligent, self-learning systems, possibly incorporating speech analysis and speech synthesis, which is the long-term solution, and the need to have systems for use at the present time.

Greater efforts must be made to persuade a wide swathe of translators to accept some form of automation, if only using the by-products of research into artificial intelligence and automatic translation.

16.1 Workstations

There is much talk of the translator's workbench (the name of a Japanese project), translator support facility (the IBM term), or translator's work station (the term used by the Canadian Secretary of State, the Digital Equipment Corporation, and Brigham Young University).

The translation services of the European Communities have adopted a programme, called SYSLING, which aims at "establishing an architecture and functional description of an integrated system of computer applications for the translation service... The ultimate goal is a system which enables the translation service to play its important role in the process of Commission document generation (from document creation through translation, modification and internal and external distribution) as efficiently as possible"..

The idea of workstations in SYSLING is that from a single console, translators being able to access word processing, text tools, ultimately specific sub-field machine translation.

The Canadian Secretary of State's office, which employs some 1,800 translators, is concerned about the future ability of the office to cope with the translation requirements of the coming decades, in view of financial restrictions on further recruitment. Even if recruitment were unrestricted, there are indications that there may not be enough language graduates emerging raw from the universities to take up the torch, as indicated in the Charpentier report, already referred to. So information processing specialists have been brought in.

The first full-scale pilot project was started in March 1988, using eight information processing translators from the Montreal office, who specialise in the translation of computer manuals. A second pilot project began in June 1988 with a five-translator team at the National Defence headquarters in Ottawa, who specialise in shipping manuals. The two projects each have dictionaries of some 40,000 terms and expressions and use Logos machine translation software. Each project immediately added an extra 6,000 terms to the 35,000 term Logos dictionaries, and another thousand were added as they arose. Translation is reckoned to take place at 1,200 words a minute. Alexing (seeking new terms to add to the dictionary (Automatic Lexicon), or Semanta, search for new semantic rules, is done as part of the process. And translators will work from SYSLING-type workstations.

16.2 Pre-editing

It seems that we shall see a considerable extension of pre-editing, to make texts suitable for machine translation. And more and more this will be done by machine, using spelling checkers, grammar checkers and the like. Once can expect that machine translation developers, after bringing themselves to admit their machines cannot tackle any form of text, will increasingly bring in their own source text checkers. One who has done this already is John Smart, whose Smart Editor is designed for use upstream of the Smart Translator machine translation system.

The successful use of Controlled language source text by Perkins Engines and the Xerox Corporation has already been mentioned. Several controlled language systems for Japanese were described by Hiroshi Uchida at the IFTT '89 conference in Tokyo.

It is important to emphasise that controlled language is not necessarily impoverished language. The examples given by Peter Pym of its use at Perkins Engines show that it makes for more cogent writing.

"We must also learn to forget awkward kinds of jargon and badly written text: it does not pay to adapt systems to bad texts. Let sloppy authors pay the price of human translation or of pre-editing", as Loll Rolling, of the Commission of the European Communities, told the 1986 Translating and the Computer conference.

"Strictly speaking, the creation of a document is not a function of the translation process. But, if this creation can be linguistically controlled by some linguistic process, automation of the rough translation becomes a lot easier", as Professor Christian Boitet said at the same conference.

Major advances in pre-editing, with consistency of terminology in machine readable form being imposed as the norm, are likely to come from a revolutionary new development in technical writing. This is known as CALS (Computer-Aided Acquisition and Logistics), which is being imposed as a standard by the United States Department of Defense. Because of the importance of the Department as a purchaser of products requiring documentation, courses have had to be organised throughout the United States, and now it has been found necessary to organise such courses in the United Kingdom. The first such UK course, organised by the Cranfield Institute of Technology, will be held in October 1989.

The launch of a new journal, *Euro Manual,* in March 1990, the first Europe-wide journal for technical writers, will help to disseminate information about such initiatives.

Loll Rolling: Machine translation: recent developments, and Christian Boitet. Current machine translation systems developed with GETA's methodology and software tools, in Translating and the Computer 8, proceedings of the 1986 Translating and the Computer conference, edited by Catriona Picken, published by Aslib, London.

16.3 Text input

Text input by manually typing it in will make any machine translation system uneconomic, and lose all the advantages of speed. It must always be remembered that the better translators can themselves type translations almost as fast, and sometimes as fast, as typists inputting text.

An example quoted by Veronica Lawson at the IFTT 89 conference in Tokyo, was that at the Nuclear Research centre in Karlsruhe, in the Federal Republic of Germany, when input was manual, a 500-page text for translation took four weeks to input; but when the Centre purchased an optical scanner, input time dropped to only three days.

Greater use needs to be made of production of source texts on to word processors compatible with the translation system, or with increased use of optical character readers, or transmission of text by electronic mail. Advances seem disappointingly slow in this area, which is a major obstacle to the spread of machine translation.

16.4 Post-editing

The general impression is that the art of post-editing machine translated text is still in its infancy.

The most advanced thinking on this seems to have occurred at the Pan-American Health Organisation, and its translations director, Dr Muriel Vasconcellos, has produced a number of papers on the subject, of which the most detailed appeared in the first issue of *Computers and Translation,* and at the 1986 Translating and the Computer conference - what one writer has called her "bag of tricks".

Essentially what is required is the development of routines, techniques, mini-routines, macros, and the sophisticated use of Search and Replace. Dr Vasconcellos has referred to it as going all the way from the reactive to the proactive, i.e. direct involvement in the machine translation process itself. The post-editors have at their disposal a routine, for example, which looks for certain known hazards, such as inversion of word order; any passage containing such a hazard is brought up automatically on to the post-editor's screen, for processing.

The US Air Force at Wright Patterson base have succeeded in achieving a partial automation of post-editing, and this was described at the 1986 World Systran conference.

Not everyone thinks in terms of automation of the post-editing process. In the Commission of the European Communities the Italian translation division in Luxembourg receive raw Systran output, which they then use as a "crib" when producing, by dictation for subsequent typing, a completely new translation!

Apart from Dr Vasconcellos, few people have given sufficient thought to the need for training people for post-editing of machine translation, which is in fact a new type of vocation.

At the Journées européennes de la traduction conference in Paris in 1987 Ian Pigott, of the Commission, touched briefly on typology, and also on the need for a new breed of post-editor, very different from the post-editor (usually called a reviser) of human translations:

"Bien que l'expérience de la typologie des documents soft encore assez limitée, il semblerait que les textes de type informatif (documents de travail, comptes-rendus de réunions, rapports d'activité dans différents domaines) se prêtent le mieux à cette application. Le volume de traduction étant très important à la Commission, les erreurs systématiques de la traduction automatique peuvent être réduites progressivement grâce aux indications que les traducteurs fournissent en retour à l'équipe chargée du développement du système.

"La post-edition est néanmoins une tâche fondamentalement différente de la traduction ou de la révision normale. Comme la plupart des erreurs rencontrées en traduction automatique n'apparaîtraient pratiquement jamais dans les traductions humaines, le post-éditeur doit déjà être très expérimenté pour pouvoir apprécier les avantages de la traduction machine. De plus, s'il veut tirer le meilleur parti des moyens disponibles, il a intérêt à post-éditer directement sur l'écran d'une machine de traitement de textes, ce qui exige également une certaine expérience."

Muriel Vasconcellos, Functional considerations in the post-editing of machine-translated output. Computers and Translation, vol. 1 (1), 1985.

Dale Bostad, Machine translation in the USAF, in Proceedings of the World Systran Conference. Luxembourg, February 1986, published by the Commission of the European Communities.

Ian Pigott, Les développements de Systran à la Commission des Communautés Européennes in Journées européennes de la traduction professionnelle, 1987, proceedings published in Encrages by the University of Paris VIII.

16.5 The training of translators

Although a number of universities and institutions of further or higher education have acquired machine translation systems, there seems to be little information available on how these machines are being used in the translation process.

There is also little concrete information on the training of translators for machine translation.

The Bureau Marcel van Dijk report gives one example, that of the Institut Supérieur de Traducteurs et Interprètes of Brussels, which during the academic year 1987 introduced courses to introduce data processing aids for translators. One can look in vain through the texts of past conference papers given on translator training at the conferences of the UK Institute of Translation and Interpreting (ITI), the Canadian Translators and Interpreters Council and the American Translators Association (ATA) without finding anything to indicate that a new breed of translator is required. At the most what is conceded is giving the translator a certain "familiarisation" with the new technology, which often seems to mean word processors.

However at the ITI 1989 conference, in his concluding address, Anthony Hartley, of the School of Cognitive and Computing Sciences at the University of Sussex, called for a broader definition of the profession, and for a different approach to training.

“Translating and interpreting are special cases of communication," he said, " founded on the existence of language barriers. The 'language industry', a concept touched on by Juan Sager, embraces other professional communicators not directly competent for mediation across languages. Translators and interpreters are necessarily dependent on the communications of others for their livelihood, and those others are increasingly reliant on linguists. How can we foster symbiosis rather than the - not always benign - parasitism that seems to characterise our current relationship with other communicators?"

The author of this paper will also be appealing for a change of approach when he speaks in the translator training session at the ATA conference in Washington in October 1989.

It is suggested that a questionnaire (see Appendix II) be sent to universities and institutions of further and higher education in the Community, in Switzerland, United States and Canada in order to elicit further information, and possibly to stimulate a more forward-looking approach.

Bureau Marcel van Dijk, Etude des Strategies de Financement et de Commercialisation de la Traduction Automatique, report for the Commission of the European Communities, June 1988.

16.6 Terminology.

There is a continual need to find translation equivalents for new terms, and to increase co-operation between translators and terminologists who may find themselves coining new equivalences simultaneously. Professor Alan Melby has addressed himself to this problem on a number of occasions.

Massive efforts are needed in both terminology and lexicography.

Often the quality of machine translation depends on the quality of the dictionary. The Electronic Dictionary Project in Japan has been started with this idea in mind. The project began in April 1986, with the establishment of the Japan Electronic Dictionary Research Institute, with joint funding from Japanese government sources and from eight major computer companies. The research and development phases is expected to last nine years, and the budget is of the order of 14 billion yen.

In Europe we have the long-established EURODICAUTOM. TEAM, also long-established, was developed by Siemens as a large-scale electronic dictionary of technical terms. The translation bureau of the government of the Federal Republic of Germany is developing LEXIS aa a large-scale electronic dictionary for technical terms.

Canada has TERMIUM, with 900,000 terms (English and French), not to mention the Banque de Terminologie de Québec.

A major development is likely to be the growing availability of such large databases, and lexicographical works such as the Oxford English Dictionary, on CD-ROM, though quite how these will be married into machine translation developments is not yet dear.

16.7 Use made of translation

One can foresee *a* considerable increase in the use made of raw translation for information scanning purposes.

Indeed, up to the present ideas about translation, including machine translation, have been consciously or sub-consciously governed by the idea that a translation has to be a "full" translation. In future we are likely to see a much more disparate approach.

At the 1986 Translating and the Computer conference Professor Juan Sager made a strong case for the typology of texts for translation and of translation activity:

"The very existence of translation machines and machine aids to translation has forced us to undertake a systematic differentiation of text types to be translated, of translation techniques - be they human or computational - and of the end products of these new and diverse systems.

"By deciding what is and what is not a suitable text for one or the other machine aid, we are faced with the necessity of undertaking a translation-oriented text-type analysis which will oblige us to rethink many theoretical positions in pragmatics and linguistics.

"By accepting that there are alternative techniques of machine assistance, we are also forced to re-assess the human processes involved in translation - how else can we achieve an optimal interaction between translator and machine?

"By realising that the end user of translations accepts a diversified end-product, which may be purely machine-produced or variously assisted, we can derive pragmatic criteria for evaluating translations in terms of the requested or intended effect, which in turn will permit us to define a set of specifications for a variety of end products."

Juan Sager, The introduction of machine aids into the translation process: an introduction in Translating and the Computer 8 (ed. Catriona Picken), proceedings of the 1986 Translating and the Computer conference, published by Aslib, London.

16.8 Interfacing

It looks inevitable that translation will become more and more integrated with other document processing stages, as this report has stated more than once. Such a development is essential if machine translation is going to enjoy the growth it ought to have.

"Users become very impatient with the slowness of language services, and may find machine translation an attractive alternative. The job that now takes weeks to get done could be fed into a computer, which would return it neatly typed in the output language in a few minutes. The fact is that as long as language services are looked upon as a self-contained system, it may be difficult to make effective use of machine translation, since adaptations to the procedures at the input and output interfaces will be decisive for the acceptability of results. If an effort is made to graft machine translation directly on to traditional manual procedures, the prospects for success are at best uncertain", said Jean Datta, speaking in 1986 on the introduction of machine translation in a large organisation.

She calls for three things, getting the input interface under control, tailoring the output interface to users' real needs, and building computer literacy among language staff.

It is important to make sure there are no more Gigatext fiascos, which set the cause of machine translation back. Jean Datta again: ...the first applications Introduced, while being simple to use, should at the same time be looked upon by the recipients of the system as being evidently useful. If it is difficult for the users to perceive that an application helps them, then they will resist its introduction, and the general progress of computerisation will be seriously impeded." This means that ft is important to give some guidance in post-editing to the translators, and in what to expect to the users.

Jean Datta, Machine translation in a large organisation, in Translating and the Computer 8 (ed. Catriona Picken), proceedings of the 1986 Translating and the Computer conference, published by Aslib, London.

SECTION 17 Promotion of machine translation

17.1 New applications

In the first 40 years of machine translation development there has been a tendency, because of the sheer fascination of the development work, to develop machine translation systems and then find a use for them. The fact that there has been at times a perceived need, as referred to earlier, has helped maintain the momentum of the development work, rather than provide close guidance as to how that work should be carried out. It might be advantageous to look out for applications first, and then see if it is possible to match machine translation to the situation? European weather forecasting, in the manner of the Canadian METEO system, or European job vacancy notices, as suggested in the section devoted to Smart systems, are two obvious suggestions.

17.2 Sensitising European business

It might be possible to run conferences to help businessmen prepare for a multilingual Europe. As languages may not be a sufficient attraction on their own, it is probably desirable to include the whole range of document and information processing, such as voice recognition, graphics scanning, patents scanning, database scanning, trade press scanning.

17.3 Sensitising the translation world

It is important to change the way translation students are recruited and trained, as indicated above.

It is important to persuade translators' organisations to accept the challenge of at least discussing the implications of machine translation, and the future role of the translator, not as an "add-on" aspect to their activities, but as a core subject.

17.4 Reference source

There is an urgent need for a reference source where people can get reliable information on machine translation, even if it is only the addresses of suppliers. Everyone involved in this business, from editors of language magazines to machine translation personnel at the Commission of the European Communities is now being pestered for information. It is difficult to know how a reference source could be established and publicised, but this ought to be discussed further. Perhaps the International Association of Machine Translation tentatively suggested at the IFTT '89 conference in Tokyo, and at the Machine Translation Summit II in Munich, could have a secretariat. The same secretariat could monitor new publications on machine translation, create a bibliography of machine translation, and act as a source of press information on machine translation.

17.5 Survey

The language industries survey, now in progress, and being conducted by *LT Electric Word* and Ink International for the European Commission, should be an on-going operation; and a directory of language industry services should be produced.

17.6 Out-reach

The world of machine translation, and of translation generally, must be encouraged to reach out to related fields, such as information science, technical writing and document production. Someone should take the initiative of calling a European conference embracing all these fields.

APPENDIX I

Proposal for a survey of the translation market

The first decision is whether the survey should be confined to Western Europe, the non-Communist developed world, or whether it is to be worldwide.

The greater part of translation is of written documents. There is some translation of computer software screen commands and messages, and there is sub-titling for the film and television industry, and these should not be ignored for the purposes of the survey. Essentially, however, texts for translation fall into two categories, those printed and published in book or journal form, and grey literature.

The figures for literary translation can be obtained from statistics published by Unesco. It would, however, be of interest to inquire of publishers whether more titles would be published if translation were more accessible. Information could be elicited from publishers by means of a questionnaire.

Information must also be obtained from the publishers of technical manuals or textbooks. Such publishers should be encouraged to define factors inhibiting the publication of more translations: cost, speed, availability of translation, etc.

The publishers of learned and academic journals, and journals for trades and professions, should also be approached. The Institut Pasteur in Paris earlier this year (1989) announced that it was switching from French to publication in English of three learned journals, and this caused considerable controversy. Inquiries should be made about the feasibility of cover-to-cover translation of more journals, in more languages. If translation were more accessible and less costly, it might be possible for learned journals in several countries to appear both in the home language and in English, thus reversing the undesirable trend of publication only in English.

The market for the translation of grey literature, i.e. all documentation which is circulated but which does not appear in printed book or journal form, will of course be much more difficult to ascertain. The figures may not even be known as recorded information to the companies or organisations in which such work is carried out.

It is suggested, therefore, that the information may have to be ascertained by a combination of questionnaire and personal contacts, both with those who commission work and those who carry out the work.

It should be possible to survey most organisations in the public sector, both international organisations and government and quasi-government organisations, though the number of such organisations should not be under-estimated. One problem might be with military and defence establishments, because of security reasons.

In the private sector it should be possible, first of all, to identify a number of major multinational corporations, where a similar approach could be made as with public sector organisations.

APPLICATIONS OF MACHINE TRANSLATION

The surveying of translation requirements in industrial companies, commercial firms and the liberal professions in various countries can only be done, it is suggested, on a sampling basis, though a survey of the work done by translation producers (translation companies, freelance translators working directly for companies, and in-house translators and para-translators) will provide some cross-checking of market volumes.

The usual sources of translation work in industrial and commercial companies are Technical Publications Departments, Sales and Marketing Departments, and Information Departments or Libraries. Such departments may act in ignorance of the arrangements made by other departments, and it is important to ensure that translation activity in a secondary location is not overlooked.

The sort of information we require to know might include the following.

In the case of international organisations, what type of material is produced for circulation to members of more than one language group, and in what volume; what proportion of this is at present translated, and into which languages; to what extent are materials in major languages accepted, *faute de mieux,* by members of other language groups, and the degree of dissatisfaction and/or inefficiency resulting from this; what is the rate of increase in the volume of material which it would be useful to translate.

In the case of other organisations and companies, what effort, if any, is put into the scanning of publications and grey literature in foreign languages in order to obtain useful information, and the extent to which potentially useful information from such sources is disregarded because of its inaccessibility; what effort is put into the translation of product documentation and other organisation or company documentation into the language of potential customers.

In all cases an attempt should be made to ascertain to what extent material in another language is being produced by people whose command of that language is less than fluent, and how machine translation might assist in such work.

The carrying out of the survey should be seen also as an opportunity to create an enhanced awareness of mainstream translation activity in general and machine translation possibilities in particular.

APPENDIX II

Proposed questionnaire to postgraduate courses in technical and specialised translation.

The undersigned is compiling a report on applications of machine translation for the Commission of the European Communities on applications of machine translation. As part of this report it would be useful to know to what extent, and in what way, your students are being prepared to use machine translation systems, which are likely to become the dominant factor in translation activity in the coming years. Any information which you consider confidential should be enclosed in square brackets [], in which case it will only be referred to statistically, without any reference to your institution. A digest of this part of the report to the European Communities will be prepared separately and forwarded to all respondents.

1. Name of institution, plus name of faculty or department concerned.

2. Contact person.

3. Please describe briefly the nature and scope of the translation course(s) offered, and the qualifications they lead to.

4. Is an introduction to keyboarding skills (learning to type) part of your course?

5. Is an introduction to word processing techniques and how they can be specifically applied to translation part of your course?

6. Is an introduction to accessing and searching remote databases part of your course?

7. Does your institution possess the software for any terminology aids, such as Termex, Ink Texttools, Supertex etc., to which your students are introduced?

8. Does your institution possess or have access to the software for any machine translation system, such as Alps, Weidner, Logos, Systran, PC-Translator, Tovna, etc.?

9. How is the topic of machine translation and machine-assisted translation dealt with in your course?

10. Are students specifically trained in the post-editing of machine-translated output?

INDEX

Numbers refer to pages

Index (continued)

Index (continued)

Index (continued)