

The historical development of machine translation

Submission for the degree of Doctor of Philosophy by Publication
at the University of East Anglia

by

William John Hutchins

(March 1999)

§1. Introduction

The application of computers to the task of translating from one natural language into another for the assistance of translators in their work has become in recent years a major commercial enterprise. There is now a multitude of translation software products on the market offering a wide range of systems, for mainframes, for workstations, and for personal computers; for the use of professional translators, and for casual use by the general public; for the production of high-quality documents, for the production of low-quality 'gists' (for information analysis or review), for translating electronic mail, for translating Web pages, and so forth. The research for improved systems continues at a high level, not just for better translation 'engines' as such, but also for newer areas of application: spoken language, information extraction, summarization, multilingual authoring, etc. With the proliferation of systems and products there has always been a need for surveys and reviews of the state of the art, the directions of research, the products on the market, and for historical overviews of the field as a whole. Active researchers and developers do not have the time or the impartiality required to write them, but they always appreciate the results. My publications in the area of what is broadly called 'machine translation' have been devoted to the provision of such reviews, surveys and historical treatments. Equally important have been publications introducing the basic principles and methodologies in the field - which, in so far as they are summaries of the contemporary situation, represent sources of 'historical' interest.

In this assessment of my personal contributions I have provided first a general brief outline of the historical development of the field (§2), my own background (§3), and a description in chronological sequence (§4) of my publications selected for this submission (out of a total of over 55 in this field). In a final section (§5) I have attempted to draw out some of the major themes that I have identified in the historical development of machine translation.

§2. Phases in the historical development of machine translation systems and computer-based translation tools

§2.1. Before the computer

It is possible to trace ideas about mechanizing translation processes back to the seventeenth century, but realistic possibilities came only in the 20th century. In the mid 1930s, a French-Armenian Georges Artsrouni and a Russian Petr Troyanskii applied for patents for 'translating machines'. Of the two, Troyanskii's was the more significant, proposing not only a method for an automatic bilingual dictionary, but also a scheme for coding interlingual grammatical roles (based on Esperanto) and an outline of how analysis and synthesis might work. However, Troyanskii's ideas were not known about until the end of the 1950s. Before then, the electronic computer had been born.

§2.2. The pioneers, 1947-1954

Soon after the first appearance of 'electronic calculators' research began on using computers as aids for translating natural languages. The beginning may be dated to a letter in 1947 from Warren Weaver of the Rockefeller Foundation to cyberneticist Norbert Wiener. The major stimulus was Weaver's memorandum in July 1949, proposing forward possible lines of research. His optimism stemmed from the wartime success in code breaking, from developments by Claude Shannon in information theory and from speculations about universal principles underlying natural languages, "the common base

of human communication”. Within a few years research had begun at many US universities, and in 1954 the first public demonstration of the feasibility of machine translation (MT) was given (a collaboration by IBM and Georgetown University). Although using a very restricted vocabulary and grammar it was sufficiently impressive to stimulate massive funding of MT in the United States and to inspire the establishment of MT projects throughout the world.

§2.3. The decade of optimism. 1954-1966

The earliest systems consisted primarily of large bilingual dictionaries where entries for words of the source language (SL) gave one or more equivalents in the target language (TL) and some rules for producing the correct word order in the output. It was soon recognised that specific dictionary-driven rules for syntactic ordering were too complex and increasingly ad hoc; the need for more systematic methods of syntactic analysis became evident. A number of projects were inspired by contemporary developments in linguistics, particularly in models of formal grammar (generative-transformational, dependency, and stratificational). They seemed to offer the prospect of greatly improved translation.

Optimism remained at a high level for the first decade of MT research, with many predictions of imminent “breakthroughs”, but disillusion grew as researchers encountered “semantic barriers” for which they saw no straightforward solutions. There were some operational systems - the Mark II system (developed by IBM and Washington University) installed at the USAF Foreign Technology Division, and the Georgetown University system at the US Atomic Energy Authority and at Euratom in Italy - but the quality of output was disappointing (although satisfying many recipients' needs for information). By 1964, the US government sponsors had become increasingly concerned at the lack of progress; they set up the Automatic Language Processing Advisory Committee (ALPAC), which concluded in a famous 1966 report that MT was slower, less accurate and twice as expensive as human translation and that “there is no immediate or predictable prospect of useful machine translation.” It saw no need for farther investment in MT research; and instead it recommended the development of machine aids for translators, such as automatic dictionaries, and continued support in basic research in computational linguistics.

§2.4. The aftermath of the ALPAC report, 1966-1980

Although widely condemned as narrow, biased and short-sighted, the ALPAC report brought a virtual end to MT research in the United States for over a decade and it had great impact elsewhere in the Soviet Union and in Europe. However, research did continue in Canada, in France and in Germany. Within a few years Peter Toma, one of the members of the Georgetown University project, had developed Systran for operational use by the USAF (1970), and shortly afterwards his system was installed by the Commission of the European Communities for translating from English into French (1976) and later between other Community languages. At the same time, another successful operational system appeared in Canada, the Meteo system for translating weather reports, developed at Montreal University.

In the 1960s in the US and the Soviet Union MT activity had concentrated on Russian-English and English-Russian translation of scientific and technical documents for a relatively small number of potential users, who would accept the crude unrevised output for the sake of rapid access to information. Since the mid-1970s the demand for MT has come from quite different sources with different needs and different languages. The administrative and commercial demands of multilingual communities and multinational trade have stimulated the demand for translation in Europe, Canada and Japan beyond the capacity of the traditional translation services. The demand was now for cost-effective machine-aided translation systems that can deal with commercial and technical documentation in the principal languages of international commerce.

§2.5. *The early 1980s.*

The 1980s witnessed the emergence of a variety of system types and from a widening number of countries. First there were a number of mainframe systems, whose use continues to the present day. Apart from Systran, now operating in many pairs of languages, there was Logos for German-English translation and for English-French in Canada; the systems for Spanish-English and English-Spanish translation developed internally at the Pan American Health Organization; and the Metal system from Siemens, initially for German-English translation and later for other languages. In addition, major systems for English-Japanese and Japanese-English translation came from Japanese computer companies, Fujitsu, Hitachi and Toshiba.

With the wide availability of microcomputers and of text-processing software came a commercial market for cheaper MT systems, exploited in North America and Europe by companies such as ALPS, Weidner, Linguistic Products, and Globalink, and by many Japanese companies, e.g. Sharp, NEC, Oki, Mitsubishi, Sanyo. Other microcomputer-based systems appeared from China, Taiwan, Korea, Eastern Europe, the Soviet Union, etc.

Throughout the 1980s research on more advanced methods and techniques continued. For most of the decade, the dominant strategy was that of ‘indirect’ translation via intermediary representations, sometimes interlingual in nature, involving semantic as well as morphological and syntactic analysis and sometimes non-linguistic ‘knowledge bases’. The most notable projects of the period were the GETA-Ariane (Grenoble), SUSY (Saarbrücken), Mu (Kyoto), DLT (Utrecht), Rosetta (Eindhoven), the knowledge-based project at Carnegie-Mellon University (Pittsburgh), and the two ambitious international multilingual projects: Eurotra, supported by the European Communities, involving teams in each member country; and the Japanese CICC project with participants in China, Indonesia and Thailand.

§2.7. *The early 1990s*

The end of the decade was a major turning point. Firstly, a group from IBM published the results of experiments on a system (Candide) based purely on statistical methods, encouraged by the success of newer stochastic techniques in speech recognition. Secondly and at the same time, certain Japanese groups began to use methods based on corpora of translation examples, i.e. using the approach now called ‘example-based’ translation. In both approaches the distinctive feature was that no syntactic or semantic rules are used in the analysis of texts or in the selection of lexical equivalents; both approaches differed from earlier ‘rule-based’ methods in the exploitation of large text corpora.

A third innovation of the period was the start of research on speech translation, involving the integration of speech recognition, speech synthesis, and translation modules - the latter mixing rule-based and corpus-based approaches. Inevitably, the subject domains have been highly restricted. The major projects are at ATR (Nara, Japan), the collaborative JANUS project (ATR, Carnegie-Mellon University and the University of Karlsruhe), and in Germany the government-funded Verbmobil project. However, traditional rule-based projects continued, e.g. the CAT2 system (a by-product of Eurotra) at Saarbrücken, the Catalyst project at Carnegie-Mellon University for the Caterpillar company, the project at the University of Maryland based on the linguistic theory of ‘principles and parameters’, and the ARPA-funded research (Pangloss) at three US universities.

Another feature of the early 1990s was the changing location of research activity. Research in the United States saw a marked revival, with the first government funded projects since the ALPAC report; in Western Europe, the main focus turned from MT as such to the development of translator workstations for professional translators, to work on controlled language and domain-restricted systems, and to the application of translation components in multilingual information systems.

§2.8. *The late 1990s.*

These trends have continued into the later 1990s. In particular, the use of MT and translation aids (translator workstations) by large corporations has grown rapidly - a particularly impressive increase is seen in the area of software localisation (i.e. the adaptation and translation of computer programs and their accompanying documentation for new markets). There has been a huge growth in sales of MT software for personal computers (primarily by non-translators) and even more significantly, the growing availability of MT from on-line networked services. The demand has been met not just by new systems but also by 'downsized' and improved versions of previous mainframe systems. While in these applications, the need may be for reasonably good quality translation (particularly if the results are intended for publication), there has been even more rapid growth of automatic translation for direct Internet applications (electronic mail, Web pages, etc.), where the need is for fast real-time response with less importance attached to quality. The services provided by CompuServe, Altavista, etc. would not have been possible at all without MT. With these developments, MT software is becoming a mass-market product, as familiar to the general public as word processing and desktop publishing.

§3. **Personal background**

After graduating from the University of Nottingham in 1960, I began my career in librarianship (a diploma at UCL, posts at the universities of Durham, Sheffield, and finally UEA.) Although my B.A. degree had been in modern languages, I became interested in linguistics only in about 1963 or 1964 when I started to read the works of Chomsky and other transformational grammarians. It became clear to me that the formalization of linguistics offered prospects for the computerization of library processes, in particular I became interested in the work beginning at that time in information retrieval (IR) and wrote my first paper ('Automatic document selection without indexing', *Journal of Documentation* 23(4), 1967, 273-290.) However, there was not much interest in linguistics approaches to IR - until the 1990s, the view prevailed that linguistics had little or no contribution to make; IR became almost wholly statistics based - and I was interested in the linguistic aspects (writing in 1975 *Languages of indexing and classification*). I did continue for some years (until the late 1980s) to write about summarization (where the use of linguistic analysis of some kind was obviously relevant), but by the middle 1980s my interest was almost exclusively in MT.

In the meantime, I had pursued some 'purely' linguistic research. In 1970 I published *The generation of syntactic structures from a semantic base* (Amsterdam: North-Holland), describing a formal model for linking semantic and syntactic representations. It attracted the attention of, among others, the foremost Russian theorist of MT, Igor A. Mel'chuk. This model might now be regarded in hindsight as a 'forerunner' of the kind of stratified models underlying a number of MT systems in the 1970s and 1980s. At the time, however, my intention was merely to extend the formalism of linguistics to cover sentence production in a 'performance' model rather than in a competence model. Hence the stress I gave to thematic organization of a quasi-Hallidayan kind.

§4. **Chronology of publications in machine translation**

The following is a description in rough chronological sequence of the publications listed in §6 below and which are offered for assessment in this submission. (References are to the numbers (#) given there.)

§4.1. *From 1978 to 1985.*

In 1978 I was invited by the *Journal of Documentation* to write a survey of recent developments in MT. The editor knew of my interest in the linguistic aspects of IR, and rightly surmised that I was also interested in machine translation. The request had been stimulated by developments in the European Commission that would have surprised many

people, coming just ten years after the ALPAC report (§2.3). Yet in 1976, the CEC had decided to purchase a MT system to deal with its burgeoning translation requirements. At the same time, there was news of the Meteo system in Canada that was translating automatically weather reports for public broadcast. The library and information community, which had followed MT developments closely in the 1960s, wanted to know what had changed. Why was the previously discredited MT now being taken up by the EC?

The article I wrote (#1) was both a survey of the situation in the 1970s (section §2.4 above) and an introduction to the current methods of MT and machine aids for translation (primarily terminological resources). It was the most complete survey of the state of the art at the time, and served (as I learnt in later years) as an introduction to MT for many researchers and students throughout the world.

The article was not intended to give an historical perspective - there were only a few passing references to the period before ALPAC (mainly citations to previous surveys) - and so, when invited to contribute to the Aslib series of conferences a couple of years later (1982), I made my first attempt to provide some kind of historical overview of MT (#2). It covered the 1950s and 1960s fairly briefly (§2.2 and §2.3), described in greater detail systems of the early 1970s (§2.4), outlined the types of 'advanced' systems then under development (including the prospects of AI approaches), and ended with some optimistic statements about the multilingual, multinational Eurotra system, then just launched.

In my next paper I attempted to explain why despite advances in computational linguistics systems based on what were then considered old-fashioned approaches were capable of producing results as good as, and sometimes better than, more recent systems (#3). The comparison was made essentially between Systran and the advanced 'transfer' systems (Ariane and Eurotra). At roughly the same time I gave a paper (#4) at a conference in Cranfield (1984) which described in some detail the methods of syntactic and semantic analysis which were under development at the time, the early 1980s (§2.5)

§4.2. Book on the history of MT, 1986

In 1983 I was approached by a publisher to write a book on machine translation - probably as a result of my 1978 article (#1). What I offered to do was a historical survey of how MT had developed, an overview of the current situation, and some prognostications of the future. In the event, the book (#5) turned out to be a substantial history of MT, based on primary sources, and providing details of all the most significant systems that had been researched and developed up to about 1984 (i.e. covering the periods §2.1 to 2.5). These historical chapters formed the bulk of the nearly 400 page long book. The coverage of contemporary systems had a few gaps, because information about the Japanese systems then beginning development was very sparse; and the chapter on future perspectives was relatively limited. The book was immediately adopted by many researchers and by teachers as their first source for information about MT - although not intended as a textbook it was used as such by many people. Now, fifteen years later its 'textbook' function has greatly diminished (in part with the appearance of #11, see 4.4 below); however, its historical value remains - it is still the most comprehensive history of MT - indeed, it is still the *only* history of MT in book form. (Later I wrote a briefer history (#16), which brings the picture up to the middle of the 1990s, but in much less detail, and at present I am attempting to write a new book on MT 'history' covering a longer period but not with the same wealth of detail.)

§4.3. From 1986 to 1992

An outcome of the book was an invitation to give the keynote speech at the first of the biennial MT Summit conferences. In this talk (#6) I considered the future prospects for MT - which, as others have acknowledged, have been surprisingly accurate (e.g. the increasing use of unrevised MT, and the impact of global telecommunications networks).

My book on the history of MT had taken the story up to about 1984. At that time

almost nothing was known about the Japanese activity in the area - the Japanese had published virtually nothing in English or any other Western language, indeed they had published very little in Japanese before 1984. It was therefore an ideal opportunity at a conference in 1988 to provide (at the organizer's invitation) an update of my history (#8), covering the continuing activity in the West in the latter part of the 1980s (i.e. period §2.6), and a survey of MT systems already available or under development in Japan and East Asia (China, Korea, and South-East Asia). When my book was translated into Chinese (in 1993) this article was also included.

In the following year, also at a conference (#9), I provided a wider ranging historical review of the 1980s in general (i.e. periods §2.5 and 2.6). This period had seen dramatic changes: the first commercial systems, major research activity in Europe and Japan, the coming of personal computers, new directions in research (interlingua and knowledge-based systems), and the increasing use of computer-based translation tools by professional translators.

There was indeed increasing interest among professional translators for what was happening, and in 1988 I wrote a general review of the future prospects for the American Translators Association (#7). A third of this article was devoted to the impact of developments in computer-based translation aids, which were only then becoming truly affordable for independent translators. Another third considered the problems of MT and the possible future 'solutions' (including artificial intelligence and spoken language systems). The last third looked at the translator's workstations which were expected to appear imminently on the market - as indeed they did, in the next two years.

Despite rapid developments during the 1980s, major improvements in translation quality had not emerged. It was evident that the kinds of the problems encountered in MT were not familiar to the general public; in particular translators misunderstood the nature of the difficulties. It was for this reason that in 1991 I gave a talk to an audience of translators (at the Aslib conference, #10) in which I attempted to describe the basic problems of automating translation, the methods being used to tackle them, and the major obstacles to high quality MT that were likely always to be with us.

§4.4. Introductory textbook 1992

Early in 1990 I received an invitation from Academic Press to write a book on MT for them. It was clear to me then that what was required was a general introduction to MT for students and for beginning researchers. I decided to approach Harold Somers, a lecturer at UMIST, who had already a number of years' experience in teaching MT, and he accepted my invitation to collaborate. We decided that the book should be directed primarily to the needs of graduate students, and we hoped that it would be of value also to other students (at all levels) and researchers in the more general field of computational linguistics. The book (#11) is divided into two main sections. In the first nine chapters we covered the basic linguistic and computational processes involved (these chapters were written jointly, each adding to and revising the work of the other in frequent reciprocal exchanges - the only exception was the chapter I wrote on evaluation). In the remaining nine chapters we devoted a chapter each to major MT systems, systems that we believed every student ought to know about (my contributions were the chapters on Systran, Meteo, Rosetta, DLT, and the final chapter covering a number of other important systems).

The book has proved to be very popular. It appears to be used on all university courses devoted to MT and on many courses of computational linguistics. In addition, we know that many present-day researchers continue to refer to it for information about basic techniques and for details of the systems described. It has been reprinted, translated into Spanish, and Harold and I continue to receive inquiries about a second edition.

§4.5. From 1992 to 1996

By the time our book was published, it was already evident that major new developments were taking place: corpus-based methods of research (statistical and

example-based), rapid commercialization (particularly PC software), the increasing growth in the use of MT systems by multinational companies, and the first translators' workstations. In two substantial surveys of the field in 1993 and 1994 (#12, #14), I provided information about new systems under development and recently marketed, the developments in research, and the significance of the direct involvement of professional translators in the practical use of computer-aided translation tools. The changes around 1990 marked, in my view, the close of the 'revival' period (from 1976, i.e. §2.4 to 2.6) and the beginning of a 'new era' (§2.7 and 2.8).

A more detailed historical treatment of the changes that had happened were given at a conference in 1994 (#15), when I described the changes in research methodologies: from rule-based to corpus-based, from syntax-oriented methods to lexicalist' methods, from an emphasis on analysis, disambiguation and 'understanding' to greater concern with for good-quality generation and more 'colloquial' output, from general-purpose systems to domain- and user-specific systems, and concluding with a perspective on future system types. This substantial historical survey covered in fact the whole period since my book (#5) to the middle of the 1990s. It enabled me to write a comprehensive history of MT and computer-aided translation (for a collection on the history of linguistics), which I entitled 'brief' in contrast to my more voluminous book. This article (#16) covers the whole history of the field from 1947 to 1994, and still represents the most complete single source for MT history. The article was in fact a substantial revision of the historical section of an earlier general article (#13) written for the multi-volume *Encyclopedia of languages and linguistics*. As an introduction to the main problems and methods of MT, it was intended for the linguist with no knowledge of the field. (I may point out that I have written other encyclopedia articles of this type, e.g. for the *Encyclopedia of computer science* (New York, 1993), for an *Encyclopedia of translation, Chinese-English and English-Chinese* (Hong Kong, 1995), and for a forthcoming *Encyclopedia of literary translation*.)

Just as at the MT Summit in 1987 I had given my ideas of what the future might hold, at the MT Summit in 1995 I reflected on current developments from a historical perspective (#17). The purpose was to draw attention to mistaken aims (e.g. the still prevalent assumption that the ultimate goal of MT must be 'perfect' translation), the rise and fall of different methodologies (many just to use MT as a testbed for some new linguistic theory), and the areas neglected in the past and now (e.g. languages not yet covered by MT research).

§4.6. Since 1996

In recent years an increasing number of my publications has been devoted to the history of (mainly) the earliest years of MT. The most important of these publications (#20) represents the fullest chronicle yet of the 'pioneer' years of MT between 1947 and 1954 (§2.2), containing much detail and information previously unpublished and/or unfamiliar to researchers and others with an interest in MT today. A general introduction to those early years was presented at a conference which celebrated the 50th anniversary of MT (#23); and at an earlier conference in the same year, I gave an assessment of the events which took place at the first MT conference in 1952 (#21), emphasizing some of the considerable similarities of basic assumptions and aims to those of current researchers. As another example of my historical papers (in a series "From the archives... " which I wrote for *MT News International*) I marked the 30 years since the (notorious) ALPAC report (§2.4) with a summary of its background, its deliberations, its recommendations and its impact (#18). Finally, on the occasion of the 20th anniversary of the Aslib conferences I presented last November an historical survey of the proceedings of this series of influential conferences (#25), highlighting and commenting on the similarities, the differences and the considerable changes over the years (i.e. covering the periods §2.5 to 2.8)

At the same time, however, I have continued to give general surveys of the 'state of the art': for example, for a conference in Montreal I was invited to review the current

developments in MT-related research in Europe (#19), which may be characterized as a general trend away from the development of MT systems as such (i.e. as independent programs) and towards the integration of translation components in domain- and user-specific natural language applications (e.g. cross-national information services, multilingual database searching, facilities for socially and physically disadvantaged members of the community, etc.). Other examples have been a presentation of developments in MT (with particular emphasis on translation aids) at a conference of translators (#22), in part seeking to dispel some of the fears of MT and computer aids that many translators still have; and a review (#24) of developments in the use of MT and translation aids by large corporations, i.e. the effects on authors, translators and the total documentation workflow within such organizations.

Finally, there are a number of publications (mainly historical) about to appear which might be mentioned: first an article on 'The origins of the translator's workstation' (for the journal *Machine Translation*), which demonstrates that its conceptual sources stretch back over 25 years; an article on Petr Petrovich Troyanskii, which will contain substantial translated extracts from otherwise inaccessible Russian publications, and will show to what extent this pre-computer pioneer anticipated later developments. Also devoted to MT pioneers is a book I am editing (for John Benjamins, Amsterdam) with contributions by and about major figures in the early years of MT in North America, Russia and Europe, and for which I am also writing two articles and the introduction. Finally, there will appear this year a comprehensive listing of commercially available MT systems and translation support tools (*Compendium of translation software*).

§5. Some themes in my work

It is not easy to identify distinctive features in one's own publications. This is particularly true, I believe, when most papers are surveys of historical developments or reviews of the contemporary situation. In such publications, the goal has to be a true, accurate and impartial reflection of what has happened or is happening, and not to give an excessively personal or biased view. Nevertheless, I believe it is possible to identify certain strands which have not been emphasized by others in the field, and which constitute observations of general validity.

The first that may be mentioned is the observation that increasing sophistication of linguistic methods of analysis and synthesis has had relatively minor impact on the improvement of the quality of the translations produced. This was seen already in comparisons (#2) between the Systran system (based essentially on 1960s methods little influenced by linguistic theory) and the Ariane and Eurotra systems of the 1980s (based explicitly on 'sophisticated' linguistic and computational methodologies). It would appear that the compilation of large comprehensive dictionaries, particularly for specific subject domains, has much greater influence on translation quality than the methods used for syntactic and semantic analysis. The observation is borne out by developments in the 1990s: in general, systems with large dictionaries still outperform many recent, methodologically more innovative systems (e.g. #17). This is, of course, a major disappointment for MT researchers - in the 1980s (#3, etc.), I had myself expected that the Eurotra research based on the latest most advanced linguistic theories and formalisms would inevitably be a major improvement on what were seen as the 'inherently' limited earlier types of systems.

A second and related point is that one of the unfortunate features of MT research has been the regular arrival of new research paradigms proposed as the 'answer' to the problems of MT (e.g. #17). Armed with the latest theories and methods, researchers have come to MT in order to demonstrate the power of new approaches. MT lends itself well to such 'testbed' use, as, in the view of outsiders, it is an easy matter to judge whether the results from one translation program are or are not 'better' than another. Unfortunately, evaluation of MT is no easy matter what may be of no use or value to one person may be useful to another. For companies what is more pertinent is whether a system can be used cost-effectively, and this may have very little to do with the intrinsic quality of the 'raw'

MT output. More crucially, however, is the fact that throughout the history of MT the initially promising results of small-scale research systems are rarely carried through into large-scale working systems. In fact, most operating systems are amalgams of many 'theoretical' approaches - experience has shown that there can be no single 'solution' to MT; and this is likely to be so into the foreseeable future. In particular, proposals for 'hybrid' solutions have come to prominence in the last decade with the integration of corpus-based approaches and more 'traditional' rule-based methods (e.g. #12, #14, #15).

A third observation is that methods and approaches of the past are subject to revivals – often without awareness by those re-introducing them that they are repeating the past. However, revivals are never exact repetitions: such are the developments in computer science that the technical bases for programs become completely different. In these circumstances, a revival can be highly beneficial. This was undoubtedly the case with the reappearance of statistical methods in the late 1980s, after some three decades of neglect.

As a fourth recurrent observation in my contributions I can mention the concentration of MT research and MT systems on relatively few languages (e.g. #8, #9, #17). The limitation was there for obvious reasons in the early years (i.e. predominantly Russian, English and other languages of 'intelligence' interest); later the bicultural and multilingual policies of Canada and the European Union focussed on other languages; and in recent years, commercial reasons have produced their own concentration on certain 'major' languages. The result has been, however, that some languages of Eastern Europe, and most languages of Africa, of India and of South East Asia have been neglected completely; yet it may be here that the benefits of cross-language transmission of technical and scientific material could be of immense significance. Just recently, my concern in this area stimulated a discussion at the AMTA conference last November; there are no obvious solutions, but awareness of the issue has been raised among those with the financial means to do something.

Finally, in everything I have written, there have been two overriding goals. One has been to inform those within the MT community about what others have done and are doing, and to place current and past activity, basic methods and approaches, and any changes that have occurred in their relevant historical contexts. The second has been to convey to others outside the community, as honestly and as clearly as possible, what has been achieved, what might be possible in the future and what remain the essential limitations of automation in this field (e.g. #10, #22).

§6. List of publications submitted

- #1. 'Machine translation and machine-aided translation'. *Journal of Documentation* 34(2), June 1978, 119-159 (Progress in Documentation) [Reprinted in: *Translation: literary, linguistic, and philosophical perspectives*. Edited by William Frawley. (Newark: University of Delaware Press, 1984); pp. 93-149]
- #2. 'Linguistic models in machine translation'. *UEA Papers in Linguistics* 9, January 1979, pp. 29-52
- #3. 'The evolution of machine translation systems'. In: *Practical experience of machine translation: Proceedings of a conference*, London 5-6 November 1981, ed. by Veronica Lawson. Amsterdam, North-Holland Publ.Co., 1982. pp. 21-37
- #4. 'Methods of linguistic analysis in machine translation'. Paper given at International Conference on Machine Translation, Cranfield Institute of Technology, February 1984. Printed in: *Progress in machine translation: natural language and personal computers*. Ed. by Ian D.K.Kelly. Wilmslow: Sigma Press, 1989. pp. 3-35.
- #5. *Machine translation: past, present, future*. (Ellis Horwood Series in Computers and their Applications.) Chichester, Ellis Horwood, 1986. 382p. ISBN: 0-85312-788-3. {Chinese translation: *Ji chifan yi: guo chyū, shan zai, wei lai*. [Taipei], Zhi-Wen Publication Company, 1993. ISBN: 957-8759-01-0. 487pp. Includes #8]
- #6. 'Prospects in machine translation'. In: *Machine Translation Summit* [Proceedings of conference held September 17-19, 1987, Hakone Prince Hotel, Japan.] Editor-in-chief: Makoto Nagao. Tokyo: Ohmsha Ltd., 1989. pp. 7-12.

- #7. 'Future perspectives in translation technologies'. In: *Technology as translation strategy*, edited by Muriel Vasconcellos. State University of New York at Binghamton, 1988. pp. 223-240.
- #8. 'Recent developments in machine translation: a review of the last five years'. In: *New directions in machine translation: conference proceedings*, Budapest 18-19 August 1988. Editors: Dan Maxwell, Klaus Schubert, and Toon Witkam. Dordrecht: Foris Publications, 1988. pp. 7-62.
- #9. 'Out of the shadows: a retrospect of machine translation in the eighties'. Paper presented at Computer & Translation '89, Tbilisi (Georgia), November-December 1989. In: *Terminologie et Traduction* no.3, 1990, pp. 275-292
- #10. 'Why computers do not translate better'. *Aslib Proceedings* 44 (10), October 1992, p. 351-359. [Presented at: *Translating and the Computer 13: the theory and the practice of machine translation - a marriage of convenience? ...* 18-19 November 1991, CBI Conference Centre, London.]
- #11. *An introduction to machine translation* (with Harold L. Somers). London: Academic Press, 1992 {March}, xxi, 362 pp. ISBN: 0-12-362830-x. [Spanish translation: *Introduccion a la traduccion automatica*. Madrid: Visor, 1995. 480pp. ISBN: 84-7774-871-3]
- #12. 'Latest developments in machine translation technology: beginning a new era in MT research'. In: *The Fourth Machine Translation Summit: MT Summit IV*. Proceedings: International cooperation for global communication, July 20-22, 1993, Kobe, Japan. (Tokyo: AAMT, 1993), pp. 11-34.
- #13. 'Machine translation: history and general principles'. In: *The encyclopedia of languages and linguistics*. Editor-in-chief: R.E.Asher. Oxford: Pergamon Press, 1994. vol. 5, pp. 2322-2332.
- #14. 'A new era in machine translation'. *Aslib Proceedings* 47 (10), October 1995, pp. 211-219 [Paper presented at *Translating and the Computer 16, ...* 10-11 November 1994, Institution of Civil Engineers, London.]
- #15. 'Research methods and system designs in machine translation: a ten-year review, 1984-1994'. In: *Machine Translation: Ten Years On*. Proceedings of the second international conference organised by Cranfield University [and] British Computer Society,... 12-14 November 1994. Cranfield, Bedford: Cranfield University Press, 1998.16pp.
- #16. 'Machine translation: a brief history'. In: *Concise history of the language sciences: from the Sumerians to the cognitivists*. Ed. E.F.K.Koerner and R.E.Asher (Pergamon, 1995), pp. 431-445.
- #17. 'Reflections on the history and present state of machine translation'. In: *MT Summit V proceedings*, Luxembourg, July 10-13, 1995. [pp. 89-96]
- #18. 'ALPAC: the (in)famous report'. *MT News International* 14, June 1996, pp. 9-12.
- #19. 'The state of machine translation in Europe'. In: *Expanding MT horizons: proceedings of the Second Conference of the Association for Machine Translation in the Americas*, 2-5 October 1996, Montreal, Quebec, Canada, pp. 198-205
- #20. 'From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology'. *Machine Translation* 12 (3), 1997, pp. 195-252.
- #21. 'Looking back to 1952: the first MT conference'. In *TMI-97: proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation, July 23-25, 1997, St. John's College, Santa Fe, New Mexico, USA*. [Las Cruces: Computing Research Laboratory, New Mexico State University] pp. 19-30.
- #22. 'Translation technology and the translator'. *ITI Conference 11. Proceedings [of] International conference, exhibition & AGM, 8-10 May 1997*. Compiled by Catherine Greensmith & Marilyn Vandamme. (London: Institute of Translation & Interpreting, 1997), pp. 113-120. [Also: *Machine Translation Review*, issue no. 7 (April 1998), pp. 7-14]
- #23. 'First steps in mechanical translation'. In: *MT Summit VI: past, present, future*.

Proceedings, 29 October - 1 November 1997, San Diego, California. Edited by Virginia Teller and Beth Sundheim. [Washington, D.C.: Association for Machine Translation in the Americas, 1997] pp. 14-23.

- #24. 'Computer-based translation tools, terminology and documentation in the organizational workflow: a report from recent EAMT workshops'. *Proceedings of the International Conference on Professional Communication and Knowledge Transfer*, Vienna, 24-26 August 1998, vol.II: 4th Infoterm Symposium: *Terminology work and knowledge transfer — Best practice in terminology management and terminography*. (Vienna: TermNet, 1998), pp. 255-268
- #25. 'Twenty years of Translating and the Computer'. In: *Translating and the Computer 20: proceedings of the Twentieth International Conference on Translating and the Computer, 12-13 November, 1998, London*. (London: Aslib, 1998). 16pp.