W. John Hutchins

REFLECTIONS ON THE HISTORY AND PRESENT STATE OF MACHINE TRANSLATION

Early pragmatism

When machine translation was in its infancy, in the early 1950s, research was necessarily modest in its aims. It was constrained by the limitations of hardware, in particular inadequate memories and slow access to storage, and the unavailability of high-level programming languages. Even more crucially it could look to no assistance from the language experts. Syntax was a relatively neglected area of linguistic study and semantics was virtually ignored in the United States thanks to the behaviourist inclinations of the leading scholars. It was therefore not surprising that the first MT researchers turned initially to crude dictionary based approaches, i.e. predominantly word-for-word translation, and to the application of statistical methods. Warren Weaver himself, in the 1949 memorandum which effectively launched MT research, had advocated statistical methods alongside cryptography, which was soon recognised as being irrelevant, and more futuristically the investigation of universal interlinguas.

With such limitations, early researchers set out with modest aims. They knew that whatever systems they could develop would produce low quality results, and consequently they suggested the major involvement of human translators both for the pre-editing of input texts and for the post-editing of the output, and they proposed the development of controlled languages and the restriction of systems to specific domains. Above all, they proposed that MT systems could progress by the cyclical improvement of imperfect approaches, i.e. an application of the engineering feedback mechanism with which they were familiar. In this atmosphere the first demonstration systems were developed, notably the collaboration between IBM and the Georgetown University in 1954.

The outcome of these early demonstrations was, however, that the general public and potential sponsors of MT research were led to believe that good quality output was achievable within a matter of a few years. The belief was strengthened by the emergence of greatly improved computer hardware, the first programming languages, and above all by developments in syntactic analysis. It was not clear which methods would prove most successful in the long run, so US agencies were encouraged to support a large number of projects. Enthusiasm for MT spread throughout the world, and in this period from the mid 1950s to the mid 1960s many of the approaches which are still current were first put forward.

The perfectionist tendency

At this time MT research grew in ambition. It became widely assumed that the goal of MT must be the development of fully automatic systems producing high quality translations. The use of human assistance was regarded as an interim arrangement: post-editing should wither away as systems improved. The emphasis of research was therefore on the search for theories and methods for the achievement of 'perfect' translations.

There were of course dissenters from the dominant 'perfectionism'. Researchers at Georgetown University and IBM were working towards the first operational systems, and they accepted the long-term limitations of MT in the production of usable translations. More influential was the well-known dissent of Bar-Hillel. In 1960, he published a survey of MT research at the time which was highly critical of the theory-based projects, particularly those investigating interlingua approaches, and which included his demonstration of the non-feasibility of fully automatic high quality translation (FAHQT) in principle. Instead, Bar-Hillel advocated the development of systems specifically designed on the basis of what he called 'man-machine symbiosis', a view which he had first proposed nearly ten years before when MT was still in its infancy (Bar-Hillel 1951).

Nevertheless, the main thrust of research was based on the explicit or implicit assumption that the aim of MT must be fully automatic systems producing translations at least as good as those made by human translators. The current operational systems were regarded as temporary solutions to be superseded in the

near future. There was virtually no serious consideration of how 'less than perfect' MT could be used effectively and economically in practice. Even more damaging was the almost total neglect of the expertise of professional translators, who naturally became anxious and antagonistic. They foresaw the loss of their jobs, since this is what many MT researchers themselves believed was inevitable.

In these circumstances it is not surprising that the Automatic Language Processing Advisory Committee (ALPAC) set up by the US sponsors of research found that MT had failed by its own criteria, since by the mid 1960s there were clearly no fully automatic systems capable of good quality translation and there was little prospect of such systems in the near future. MT research had not looked at the economic use of existing 'less than perfect' systems, and it had disregarded the needs of translators for computer-based aids.

While the ALPAC report brought to an end many MT projects, it did not banish the public perception of MT research as essentially the search for fully automatic solutions. The subsequent history of MT is in part the story of how these is this mistaken emphasis of the early years has had to be repaired and corrected. The neglect of the translation profession has been made good eventually by the provision of translation tools and translator workstations. MT research has turned increasingly to the development of realistic practical MT systems where the necessity for human involvement at different stages of the process is fully accepted as an integral component of their design architecture. And 'pure' MT research has by and large recognised its role within the broader contexts of commercial and industrial realities.

Hence since the early 1970s development has continued in three main strands: computer-based tools for translators, operational MT systems involving human assistance in various ways, and 'pure' theoretical research towards the improvement of MT methods.

Tools for translators

It is now evident that MT as such is not appropriate for translators. They do not like subservience to a machine; they do not want to revise the poor quality of MT systems. What they want are sophisticated translation tools, e.g. translation

workstations, which can make their work more productive without taking away the intellectual challenge of translation. What professional translators need are tools to assist them to translate: access to dictionaries and terminological databanks, multilingual word processing, management of glossaries and terminology resources, input and output communication (e.g. OCR scanners, electronic transmission, high-class printing). For these reasons, the most appropriate and successful developments of the last few years have been the translator workstations.

The development of translation tools became feasible only since the 1960s, firstly with the availability of real-time interactive computer environments, then the appearance of word processing in the 1970s and of microcomputers in the 1980s and, subsequently, networking and larger storage capacities. From the computational linguistics perspective, the development of translation tools and translation workstations is not as challenging as MT itself for those with 'perfectionist' inclinations, and indeed this research has taken place primarily in non-academic environments.

With the appearance of workstations which have appreciably aided the day-to-day work of professional translators, there are clear signs that the previous antagonism of translators to the MT community in general is disappearing. These tools are seen to be the direct result of MT research. The most recent addition has been the `translation memory' facility which enables the storage of and access to existing translations for later (partial) reuse or revision or as sources of example translations; and this facility derives directly from what was initially 'pure' research on bilingual text alignment within corpus-based MT (see below).

Practical machine translation

All current commercial and operational systems and probably most future ones produce output which must be edited (revised) if it is to be of publishable quality. Only if rough translation is acceptable for information-gathering purposes can the output of MT systems be left untouched by human revisers. It follows that commercial developers of MT must provide adequate facilities for the revision of texts. They must also stress to customers that MT does not and cannot produce

translations acceptable without revision. It is a lesson which was learnt, painfully in some cases, during the 1980s and most, perhaps all, current MT developers and system vendors stress the imperfect nature of MT output.

MT systems have been developed on the assumption that they would be used primarily by bilinguals. In practice, the post-editing of MT output has been given to people knowing both the source and the target languages, i.e. most often to professional translators. It is a practice which has damaged the image of MT among translators: they do not wish to be revisers of poor quality output from a machine. In more recent years, the training of bilingual staff specifically for post-editing has been advocated and successfully implemented in many MT operations. The lesson has been learnt that MT post-editing should not be imposed upon professional translators; it is better to train people specifically for this role.

It is now widely accepted that MT works best in domain-specific and controlled environments. In this respect, MT developers have effectively taken up the themes and suggestions first propounded by the pioneers in the 1950s. Sublanguage systems were also an early proposal - in the form of microglossaries - and since the success of Meteo in the mid 1970s have remained the focus of research systems. The control of language was in the rather crude and easily dismissable form of highly simplified 'model English' (Dodd 1955). The idea was largely forgotten until the practical application of Systran by Xerox in the late 1970s. Other applications of controlled input followed with the available general-purpose systems. Now it is recognised that MT systems should be designed ab initio for controlled language, and a number of independent companies outside the academic MT research community have been doing so in recent years (e.g. Volmac). The largest current development is the Caterpillar project based on the research at Carnegie Mellon University.

One of the most important lessons from the history of MT research is that there can be no quick results. The development of operational systems is a long term commitment. Even good working prototypes with reasonably large dictionaries and many years of testing cannot be easily scaled up for fully reliable operational installation. Consequently no commercial system can incorporate the very latest

methodology and technology, and this fact should be made more clearly to potential purchasers. The emphasis should be on reliability and economic viability.

The construction and compilation of dictionaries is essential to the success of any and all MT systems. But there are no easy solutions. The adaption of conventional dictionaries to the needs of MT is not trivial. Much attention was paid to dictionaries in the 1950s and 1960s - indeed the 'direct translation' approach required most selectional and structural information to be accessed from lexical entries. In the years following, problems of syntactic analysis and structural transfer dominated research and the lexicon was relatively neglected for many years. Since the mid 1980s, with the adoption of constraint-based and unification grammatical formalisms there has consequently been a more 'lexicalist' orientation, and the lexicon has become again the focus of much current MT research. It is to be hoped that this research will both accelerate the development of MT prototypes and improve the operational flexibility of commercial systems.

Research methods for machine translation

Research systems have often been developed without any idea of how they might be used or who the users might be. In many cases, researchers have begun with the intention of exploring the potential of a single theory, method or technique. It has often been the case that possible practical use is considered only after a prototype system has been built and evaluated and its operational limitations have been defined. It can be argued that the relationship should if anything be reversed, with research as the hand-maiden of practical MT.

What is forgotten by many is that MT is not a theoretical science; it is the application of computational, linguistic, etc. methods and techniques to a practical task. Translation is itself a means to an end: the communication of a message or information in a language other than that it was originally composed. It is a task which has never been and cannot be 'perfect'; there are always the possibilities of multiple translations of the same text or message according to different circumstances and requirements. MT is no different: there cannot be a 'perfect' automatic translation. The use of an MT system is contigent upon its cost effectiveness in practical situations.

Nevertheless, MT research continues to attract the perfectionists. It has been regarded as a field in which new linguistic formalisms or new computational techniques can be tried out: MT has been seen as a testbed for theories. The reason is obvious: the quality of MT and translation can be judged by non-experts, at least in a superficial manner - reliable and systematic evaluation is quite another matter.

The list of such applications of 'external' theories is long. It began in the 1950s and 1960s with information theory, categorial grammar, transformational-generative grammar, dependency grammar, and stratificational grammar. In the 1970s and 1980s came MT research based on artificial intelligence, non-linguistic knowledge bases, formalisms such as Lexical-Functional Grammar, Generalized Phrase Structure Grammar, Head-driven Phrase Structure Grammar, Definite Clause Grammar, Principles and Parameters, Montague semantics. In the 1990s have been added neural networks, connectionism, parallel processing, and statistical methods, and many more.

In nearly every case, it has been found that the 'pure' adoption of the new theory was not as successful as initial trials on small samples appeared to demonstrate. Inevitably the theory had to be adapted to the demands of MT and translation, and in the process it became modified. But innovativeness and idealism must not to be discouraged in a field such as MT where the major problems are so great and all promising approaches must be examined closely. Unfortunately, there has been a tendency throughout the history of MT for the advocates of new approaches to exaggerate their contribution. Many new approaches have been proclaimed as definitive solutions on the basis of small-scale demonstrations with limited vocabulary and limited sentence structures. It is these initial untested claims that must always be treated with great caution. This lesson has been learnt by most MT researchers; no longer do they proclaim imminent breakthroughs.

The history of MT research has gone through a number of phases in which certain frameworks have dominated. From the late 1960s the syntactic orientation was dominant, initially with syntactic transfer approaches (e.g. at MIT), then the interlingua formalisms of CETA and LRC, followed by the "second generation" transfer-based multi-level model of GETA-Ariane, SUSY, Mu, and Eurotra. In the 1980s the AI orientation was popular (e.g. Carnegie Mellon), more attention was

paid to semantics and interlingua-based systems were explored (e.g. Rosetta and DLT). And now in the 1990s, the corpus-based paradigm with stochastic and example-based methodologies is the focus of much activity.

Thus we see the rise and fall and subsequent revival of methods and approaches. Statistics-based MT disappeared for almost thirty years between the early 1960s and the IBM project Candide at the end of the 1980s. The interlingua idea has also had periods of neglect: Weaver's suggestion in 1949 was not taken up until the late 1950s by researchers in Cambridge, in Moscow and Leningrad, and when valuable theoretical research was undertaken; it flourished for a while in the syntactic interlinguas of Grenoble and Texas, and then for a decade it was considered too ambitious and the transfer-based approach was preferred until interlingua system came back again in the mid-1980s.

The lesson to be derived from such fluctuating fortunes is that no old or unfashionable theory or approach should be disregarded simply because it has once been found inadequate. MT has a long history, longer that many of those who have only recently entered the field are often aware of. Before applying some new approach to MT on a large scale, researchers and their funders should assure themselves that previous work is not about to be replicated

The advantage of this long experience is that there are many old wheels which do not have to be reinvented. There are large areas of morphological and syntactic analysis which can be adopted successfully by any new system. Indeed the success of many custom-built systems in recent years demonstrates that methods of MT and of computational linguistics are becoming widely known outside the narrow research community and can be applied with success in working MT systems.

New directions and challenges

Within the last ten years, research on spoken translation has developed into a major focus of MT activity. Of course, the idea or dream of translating the spoken word automatically was present from the beginning (Locke 1955), but it has remained a dream until now. Research projects such as those at ATR, CMU and on the Verbmobil project in Germany are ambitious. But they do not make the mistake of attempting to build all-purpose systems. The constraints and limitations

are clearly defined by definition of domains, sublanguages and categories of users. That lesson has been learnt. The potential benefits even if success is only partial are clear for all to see, and it is a reflection of the standing of MT in general and a sign that it is no longer suffering from old perceptions that such ambitious projects can receive funding.

International collaboration which appears so attractive and so obvious in MT is in fact difficult to manage. The Eurotra project was successful in the stimulation of good MT-related research in the European Communities, founding as a result the basis for future language engineering projects, but it did not produce a large-scale operational system. The CICC multinational project in Asia is still running; it too will certainly stimulate MT-related research and computer technology in Asian countries, but whether it will result in an operational system has yet to be seen.

Most MT research and virtually all commercial MT activity has concentrated on the major international languages: English, French, German, Spanish, Japanese and Russian. The languages of the less developed countries have been largely ignored. Yet it can be argued that the need for MT is as great and sometimes greater than in the more developed countries. A future challenge for the MT community must be the development of translation tools and working MT systems for these neglected languages.

Other potential areas for MT have also been neglected for many years. It is only in recent years researchers have begun investigating systems for monolinguals who are ignorant of the target language. In these cases, what is required is a means of conveying a message in an unknown language; it does not have to be a straight translation of any original. From interactive dialogue a translatable (MT-amenable) 'message' can be composed for automatic conversion into an idiomatic and correct message in the target language without further involvement of the originator. Such systems require much more attention to the generation of texts than has been customary in previous periods of MT research. As a consequence, it could well benefit the future output of basically bilingual MT systems.

Translation for those ignorant of the source language has been provided until recently from the unrevised outputs of the older batch-processing systems, i.e. as by-products of systems primarily intended to produce translations for revision before publication. Within the last decade, however, cheap PC-based software has appeared on the market which can be used by monolinguals wanting the gist of texts. In general the quality is poor, and it is unlikely that after initial delight for any assistance with translation the general public will find the quality satisfactory for much longer. Users will be expecting continuous improvements, as they do with all commercial products. The question is thus whether the manufacturers of these products prepared to invest in the research required or whether researchers in the MT community are going to address this particular need. Assumptions which are valid for systems intended for those knowing both source and target languages are not applicable in systems for monolingual users.

The recently announced versions of Systran and some Japanese systems for personal computers has undoubtedly intensified the competition in this area. As yet they are more expensive than the cheaper poor-quality products which have been already on the market for some years. It will surely not be long before the general public will want to know whether they are getting value for their money. In the past, the purchasers of MT systems were generally large companies with the resources to evaluate systems for their own particular circumstances and requirements. What is now required are standards and performance indicators which can be readily understood by potential purchasers. The MT community is now aware of this need, but the time is short and there is a risk that external evaluators will fill the gap. Undoubtedly, MT has suffered from the exaggerations during the pre-ALPAC era; expectations were raised which could not be fulfilled. There is a danger that once again there will be expectations from MT products which are unrealistic. It is incumbent upon MT researchers, developers and particularly vendors to ensure that purchasers of MT systems are not again being mislead by exaggerated claims.

In the future, much MT research will be oriented towards the development of `translation modules' to be integrated in general `office' systems, rather than the design of systems to be self-contained and independent. It is already evident that

the range of computer-based translation activities is expanding to embrace any process which results in the production or generation of texts and documents in bilingual and multilingual contexts, and it is quite possible that MT will be seen as the most significant component in the facilitation of international communication and understanding in the future `information age'.

In this respect, the development of MT systems appropriate for electronic mail is an area which ought to be explored. Those systems which are in use (e.g. DP/Translator on CompuServe) were developed for quite different purposes and circumstances. It would be wrong to assume that existing systems are completely adequate for this purpose. They were not designed for the colloquial and often ungrammatical and incomplete dialogue style of the discussion lists on networks.

When launching MT, Warren Weaver (1949) emphasised the role of translation in global communication. He began his memorandum with the statement "There is no need to do more than mention the obvious fact that a multiplicity of languages impedes cultural interchange between the peoples of the earth, and is a serious deterrent to international understanding." In the context of the international network of communication, there will still remain the barriers of language for which computer-based translation must play a major part in surmounting. Is the MT community prepared to rise to the challenge of developing systems for the global information society of the next century in which almost anyone anywhere might need translation facilities at any time for a multiplicity of documents, texts, and messages?

Bibliographical note

Details of systems mentioned and information on the historical development of MT can be found in Hutchins (1986, 1988, and 1993). Weaver's memorandum of 1949 and the 1955 articles of Locke and Dodd can be found in collection edited by Locke and Booth.

ALPAC (1966): *Language and machines: computers in translation and linguistics*. A report by theAutomatic Language Processing Advisory Committee... Washington, D.C.: National Research Council.

Bar-Hillel, Y. (1951): The state of machine translation in 1951. *American Documentation* 2, 229-237.

Bar-Hillel, Y. (1960): The present status of automatic translation of languages. *Advances in Computers* 1, 91-163.

Hutchins, W. J. (1986): *Machine translation: past, present, future*. Chichester: Ellis Horwood.

Hutchins, W. J. (1988): Recent developments in machine translation. In: Maxwell, D. et al. (eds.) *New directions in machine translation*. Conference proceedings, Budapest 18-19 August 1988. Dordrecht: Foris; 7-63.

Hutchins, W. J. (1993): Latest developments in machine translation technology. In: *MT Summit IV: International Cooperation for Global Communication, July 20-22, 1993 (Kobe,Japan)*, 11-34.

Locke, W.N. and Booth, A.D. (1955), editors: *Machine translation of languages: fourteen essays* (Cambridge, Mass.: M.I.T.Press)

[Paper presented at the MT Summit, Luxembourg, 1995]

Source : http://ourworld.compuserve.com/homepages/WJHutchins/MTS-95.htm