

PROSPECTS IN MACHINE TRANSLATION

W. John Hutchins

(University of East Anglia, Norwich, England)

Research on machine translation (MT) began in the 1950's and has largely remained to this day an activity which combines an intellectual challenge, a worthy motive and an eminently practical objective. The challenge is to produce translations as good as those made by human translators. The motive is the removal of language barriers which hinder scientific communication and international understanding. The practical objective is the development of economically viable systems to satisfy a growing demand for translations which cannot be met by traditional means. However, the realisation has been disappointing in many respects; and, although recently optimism has been growing, operational MT systems are still far from satisfactory.

No current operational MT systems can produce good quality output without either placing restrictions on input texts or involving human assistance before, during or after translation processes. Present MT systems make 'simple' grammatical errors which no human translator would make. All have difficulties in the selection of pronouns, prepositions, definite and indefinite articles, and so forth. The revision of MT output typically involves a great deal of low-level correcting, and it is not surprising that translators have a generally poor opinion of MT systems. The repetitive correction of the same mis-translation is irritating both for post-editors and for operators of interactive MT systems. Much can be done to simplify the editing facilities, e.g. for making common alterations quickly (cf. PAHO), and many of the difficulties can be avoided by limiting systems to particular subjects or styles of language. This can reduce problems of lexical homography and syntactic ambiguity, and at the same time enable higher levels of comprehensiveness and completeness in dictionaries. There is the 'sublanguage' approach (e.g. METEO), and there is the 'restricted language' approach: either systems designed for specific types of texts, such as titles and abstracts (e.g. TITRAN), or the 'regularization' of input texts (e.g. Xerox). All are means of overcoming the limitations of existing operational systems. But ultimately significant improvements can come only from fundamental research on fully automatic translation.

The dominant framework for most MT systems under current development is based essentially on the syntax-oriented approach of computational linguistics (the successor of earlier lexicon-based approaches of 'direct translation' systems). There are two basic 'indirect' strategies: interlingual and transfer. The initial projects based on the interlingua approach proved too ambitious, and in the last two decades there has been a general agreement within the MT community on the basic transfer system design: three stages of analysis, transfer and synthesis, with relatively language-independent tree-transduction procedures and abstract language-specific intermediary representations incorporating information from all levels of analysis.

The approach has its major strengths built up during many years of experience with large-scale complex systems: a solid body of well-tested and efficient methods of morphological and syntactic analysis (CFG parsers, ATN parsers, tree transducers, charts, etc.), now supported by theoretical developments in 'unification grammars' (e.g. LFG, GPSG and DCG), and the modularity and flexibility of system architectures permitting progressive incorporation of newer techniques, including recent artificial intelligence (AI) methods.

Its weaker features include the familiar limitation of analysis to sentences (with problems of pronouns, intersentential relationships, and elliptical constructions), and the constraints of the syntactic framework, However abstract and language-independent, for semantic representations. In semantic analysis there has been successful treatment of homography and syntactic ambiguity; and there have been successful implementations of case frames, of semantic features, of distributional semantic information, and recently of Montague semantics; but, nevertheless, the profounder problems of interlingual semantic analysis have proved elusive.

The abstractness of intermediary representations and therefore the depth of semantic analysis depends on whether the system is intended to be bilingual or multilingual. Bilingual systems can operate with direct equivalences of lexical items and syntactic structures, and can ignore interlingual or 'universal' considerations. By contrast, multilingual systems tend towards interlingual semantic representations (as in Eurotra), with consequential exacerbation of semantic problems; fully automatic multilingual systems based on the transfer approach are perhaps still too ambitious in the present status of linguistic and computational knowledge.

While there is undoubtedly still scope for considerable improvements in advanced transfer approaches, greater expectations are focused on AI approaches. In this framework, translation involves the 'understanding' of source texts and the expression of the (language-independent) 'meaning' in a target language text. No rigid distinction is made between linguistic and non-linguistic information, nor between syntax and semantics; text processing is not limited to sentences; understanding and text interpretation refer to knowledge databases and inference mechanisms of specific subworlds.

Most AI methods have been developed in small-scale projects. While there are no large-scale purely AI-based translation projects at present, many 'linguistics'-based MT research systems are now introducing AI techniques in various forms: greater integration of syntactic and semantic information, the incorporation of 'subworld' knowledge databases, and the embodiment of translators' expertise in disambiguation and transfer components ('expert systems'). But there is still scepticism about their general applicability. The first issue is the huge size of AI knowledge bases for fragments of natural language vocabulary much smaller than those required in MT systems even when restricted to specific sublanguages. The second issue is the doubt whether 'comprehension' to the depth assumed by AI researchers is really necessary for translation purposes. A third issue is the abstractness of 'content' representations resulting in losses of information about 'surface' structures of texts; versions produced by AI methods are not translations but rather paraphrases.

What is generally agreed is the future of MT as one component of integrated man-machine communication systems which include authoring, summarization, electronic mail, word processing, publishing facilities, access

to remote databases, and so forth. One sign already is the interest of many MT projects in translation workstations. To this picture must be added the obvious attractions of spoken language input and output, i.e. translation systems for telephone communication and ultimately simultaneous interpretation. The prospects are still distant, however: while speech production is fairly well advanced, the problems of speech recognition are much more intractable.

It would appear that the best prospects for future fully automatic translation systems will be those combining the virtues of the traditional 'linguistics' approaches and the newer 'knowledge-based' approaches. The most likely option is the integration of AI techniques within the well-founded 'computational linguistics' framework of powerful and flexible transfer systems. Advances in microcomputer technology have made possible a proliferation of small-scale MT projects, and there will be many AI-inspired projects, but major advances will probably come primarily from the larger projects (Grenoble, Eurotra, Kyoto, etc.) Only these will have the capacity to test new linguistic and AI techniques on a sufficiently large scale.

Experimental projects can be innovatory, but systems designed to be operational within a fixed timespan must be founded on techniques that ensure reasonable success. MT systems have long gestation periods; the commercial risks are considerable, with high investment in system development and in after-sales maintenance; and so there is a natural reluctance to invest in untried experimental methods. Basic MT research will continue to be primarily pursued at academic institutions, with increasing collaboration of commercial and industrial interests. International coordination is highly desirable, particularly in dictionary construction; there is too much duplication of MT research effort. The complexity of MT computational processes demands close collaboration between the research fields of linguistics, artificial intelligence and computer programming. The separation of tasks in MT projects between linguists and programmers has already become less meaningful in recent developments. Close cooperation will be even more vital when the promised era of parallel processing comes.

In the immediate future, operational MT systems designed for the production of quality translations will continue to have four options: the use of 'regularized' or pre-edited input, the restriction of systems to specific sublanguages, the involvement of human interaction during machine translation processes, and the acceptance of the need for more or less extensive post-editing of texts. The options are not mutually exclusive since post-editing may be needed for restricted language systems, and interactive systems may be limited to particular sublanguages.

For the foreseeable future, the 'regularized input' option will be economically viable only in multilingual situations, where the same text is to be translated into a number of languages. This does not mean that the MT system itself has to be multilingual; there can be (as in the Xerox case) a series of bilingual systems. Interactive MT systems are in principle feasible for simultaneous multilingual output, and this is the intention of the Utrecht DLT project. For the more traditional 'batch processed' MT with post-editing there will continue to be bilingual systems, with the newer transfer systems promising better quality than those operating at present. There is an undeniable need for multilingual systems, but it seems likely that the bilingual design will still be popular simply because multilingual systems have an inherently greater complexity and therefore greater risk of failure.

The attitude of translators to MT is unlikely to change dramatically.

Some will be prepared to work as post-editors and others will operate interactive systems. But most will prefer sophisticated machine aids, accessed at a single terminal, e.g. a translator's workstation, combining aids which are already becoming commonplace: on-line text editing, in-house specialized glossaries, on-line access to remote term banks, transmission of texts by telecommunication links, optical character recognition, and production of high-quality printed output. In this way they can achieve higher rates of productivity and greater levels of consistency while maintaining their traditional high standards, particularly in scientific, technical and administrative translation. There is still room for much improvement; irritating equipment incompatibilities, inadequate provision of non-English characters and non-Roman alphabets, etc., but there are solutions within existing technologies and more advanced facilities will come from MT and AI lexicographic and semantic research. The growth of MT has in fact increased the options for translators; they will be able to concentrate on translations which need their creative flair, and they will leave the tedious, repetitive and mundane to the machine.

The area of translation with greatest potential for MT systems is the largely untapped demand for rough translations. The information needs of scientists, technicians, and administrators can often be satisfied by the unedited output of existing MT systems. The recipients must have the subject knowledge enabling them to understand the intentions of authors even if the message is obscured by lexical and grammatical errors and by stylistic idiosyncrasies. The toleration of errors varies, however, according to the nature of the material translated; greater accuracy is probably required for titles and abstracts than for full texts, where the larger context provides more clues for interpretation. Even greater tolerance may be found if recipients also have some knowledge of the source language; for these virtually word-for-word 'pidgin' versions may be comprehensible. In general, however, the aim must be much higher quality than in present systems, since as unedited output improves so demand will increase and even higher expectations will be voiced. But what is not necessarily required is an AI-type 'understanding' component, since rough translations are intended for those who can call upon their own far richer knowledge resources. It is clear that in the immediate future, the best prospects lie in systems designed for particular sublanguages. Most demand will come from the applied sciences, engineering, and social sciences, where there tends to be less knowledge of foreign languages than in the pure sciences.

The most innovative direction for future MT development exploits the recognition that AI systems do not produce strict translations of a text but rather paraphrases in another language. The idea is that 'paraphrase translations' should be produced by systems which combine composition in the user's own language and simultaneous translation into another. In this case, the most immediate application is clearly conventional business correspondence, where what is most important is that the basic 'message' is conveyed in the foreign language. If speech recognition and spoken output were also to be developed within the same restricted area, the commercial prospects would look very attractive and there would be no lack of financial support.

Paraphrase translation in the other direction, into the mother tongue of the recipient, would also be a feasible interactive MT model. There are undoubtedly a sufficient number of specialists with some basic knowledge of a foreign language who would be capable of translating a text in their subject field with minimal computer assistance. Developments in this direction might eventually, with progressive miniaturization, lead to the long predicted pocket 'personal translators' for travellers and businessmen.

In the more distant future there must be the prospect of systems combining translation and summarization. The prospect of producing summaries of foreign language reports or articles for administrators, businessmen and scientists in their own language is almost certainly more attractive than paraphrase or rough translations of full texts. AI researchers and others have begun experiments on small-scale summarization programs in restricted subworlds, but it has already become apparent that the complexities of the task are at least equal to those of MT itself.

A major factor to be considered in the development of any future MT system will be its capacity to be integrated easily into many kinds of computerized office environments. This entails not just equipment compatibility, input and output facilities (OCR, network links, printing and publishing, access to remote databanks, etc.), facilities for creating in-house glossaries, easy system maintenance, but above all absolute reliability of hardware and software.

Reliability and robustness are essential. Systems must not fail either because of the unforeseen effects of dictionary changes or because of ungrammatical or illogical input texts. The levels of tolerance which may be observed at present will decrease as the quality of MT output improves and as MT is integrated more closely with other computer and telecommunication systems. Least tolerance is to be expected from businessmen using 'para-translation' systems and from translation agencies operating 'batch processing' and interactive MT systems.

Whatever the MT system, the quality of dictionaries is crucial. Good dictionaries require many years of painstaking work; arguably the present success of Systran is attributable primarily to the amount of effort which has been devoted to its large and detailed dictionaries. It is vital that operators who are not familiar with computers and programming should be able to amend and update dictionaries easily. It is also obviously desirable that MT dictionaries should be more compatible with each other, so that lexical information from one can be used in another.

The progressive improvement of MT systems, desirable for many reasons (not least to reduce post-editing costs), argues for the maximum flexibility and modularity of system design and in particular for good communication between designers, users and translators. A failure of past MT research has been the neglect of the expertise of professional translators, lexicographers and terminologists. MT systems must be built to respond to the actual experience of users, to provide facilities which users perceive as necessary for improving performance.

Predictions in the MT field have in the past been notorious for excessive optimism or for disillusioned pessimism. The safest prediction is that there will be a long-term future for nearly all varieties of systems. Machine aids, workstations, and interactive MT will be preferred by professional translators. Post-edited MT will continue in large translation services and in translation bureaux. Restricted input MT will remain the option for some multinational companies, unless 'raw' MT output improves greatly. Unrevised MT and 'para-translation' systems will be adopted increasingly for information-only translations. Automatic telephone translation seems unlikely in the near future, but universal access on demand via public networks to many different kinds of automatic translation facilities, whether interactive or 'para-translation' or unrevised 'batch' systems, does not now seem to be in the least Utopian. As always, it is the responsibility of the MT community to ensure that the general public is not disappointed by unrealistic promises.