[From: *Perspectives in artificial intelligence, vol II*, ed. J.A.Campbell and J.Cuena (Chichester : Ellis Horwood, 1989)]

Machine translation: practical issues

J. Slocum, Microelectronics and Computer Technology Corporation (MCC), 3500 West Balcones Center Drive, Austin, Texas 78759, USA

SUMMARY

Many of the problems with (and consequently much of the resistance to) machine translation (MT) has sprung from ignorance: on the part of early developers, who failed to appreciate the magnitude of the problems they faced, and hence made outrageous claims; on the part of opponents, who frequently fail to understand what translation is about, and how it is (properly) performed; on the part of zealots for theoretical formalisms, who seldom if ever try to produce working systems and consequently lack appreciation of the magnitude of the practical problems to be faced; on the part of modern developers, too many of whom know little about translators' needs and other human engineering concerns; and on the part of translators, who often fail to realize the potential for improvement in the quality of their working lives when computers become able to take over much of the 'grunt work' in translation.

In this chapter we confront these and other problems. We characterize areas where knowledge has been lacking, and discuss the consequences of such ignorance; provide background knowledge enabling the reader to acquire a more balanced perspective from which to judge the developments in the field; and finally present the author's opinions, based upon his experience as an MT system developer, regarding some of the most important issues in the construction of a practical system. This will entail, in places, an exposition of certain aspects of the underlying technology; however, complete understanding of these technical matters is not a pre-requisite for appreciation of the major points.

1. A BRIEF HISTORY OF MT

Translation of human languages was one of the first applications considered for digital computers. Indeed, the idea of mechanizing translation predated the invention of such machines, but it was only after World War II, when digital computers became generally available, that MT was taken up seriously, as a consequence of a memo by Warren Weaver. During its first decade, in the 1950s, interest and support was fueled by visions of high-speed, high-quality translation of arbitrary texts (especially those of interest to the military and intelligence communities, who funded MT projects quite heavily), with no need for human intervention. During its second decade, in the 1960s, disillusionment crept in as the number and difficulty of the linguistic problems became increasingly obvious, and as it was realized that the translation problem was not nearly so amenable to fully automated solution as had been thought. The climax came with the delivery of the US National Academy of Sciences ALPAC report in 1966, condemning the field and, indirectly, its workers alike. The ALPAC report was criticized as narrow, biased, and short-sighted but, nevertheless, it resulted in the cancellation of MT projects in the US and, though to a much lesser extent, elsewhere around the world.

By 1973, the early part of the third decade of MT, only three government-funded projects were left in the US, and by 1976 there were none. Paradoxically, MT systems were still being used by various government agencies in the US and abroad, because there was simply no alternative means of gathering information from foreign (Russian) sources so quickly. Private companies were also developing and selling (mostly outside the US) MT systems based on the mid-1960s technology so soundly condemned by ALPAC. However, the general disrepute of MT resulted in a rather quiet third decade.

We are now well into the fourth decade of MT, and there is a resurgence of interest throughout the world—plus a growing number of MT and MAT (machine-aided translation) systems in use by governments, business and industry. This year, as much as a million pages of text will be translated by machine, worldwide. Industrial firms are also funding M(A)T research and development projects of their own; thus it can no longer be said that only government funding keeps the field alive. Indeed, in the US there is virtually no government funding, though the Japanese and European governments are heavily subsidizing MT research and development.

2. THE MT CONTROVERSY

Machine translation of human languages is not a subject about which many scholars feel neutral. This field has had a long, colourful career, and boasts no shortage of vociferous detractors and proponents alike. In this section, we attempt to clarify the misunderstandings responsible for so much contention, and provide an initial basis for a more informed assessment. We close with a discussion of some of the psychosocial issues affecting acceptance of MT systems, and a summary of some lessons to be learned.

2.1 Early researchers' ignorance

Early MT researchers proved to be victims of their own enthusiasm and optimism. They had been convinced that producing general translation systems, able to deal with a wide range of texts without significant human intervention, and to produce high quality translation as output, would soon result, given nothing more than hard work. The 1954 Machine Translation Conference at Georgetown University had finished on a very optimistic note: 'mechanical translation was not only feasible but far closer to realizations than possibly the audience recognized' (Reynolds. 1954). This was not an isolated view: 'In about two years [from August 1957], we shall have a device which will at one glance read a whole page and feed what it has read into a tape recorder and thus remove all human co-operation on the input side of the translation machines. ... it will not be very long before the remaining linguistic problems in machine translation will be solved for a number of important languages' (Reifler, 1958). Obviously, it was easy for the ALPAC committee to conclude that the results achieved did not fulfil the promises made.

2.2 US critics' ignorance

In the US — a large, essentially monolingual society—it is still common to conceive of translation as simply that which any human translator does. It is generally believed that a few years' study of a foreign language qualifies one to be a translator, for just about any material whatsoever. Native speakers of foreign languages are therefore considered to be even better qualified. (In June 1986, a TV commentator, discussing President Reagan's proposed Workfare program, listed as typical positions to be filled by the otherwise-unemployed: 'clerks, street cleaners, and translators'.) Worse, fluency in one foreign language is sometimes taken to be, *ipso facto,* proof of competence in related languages; an embarrassing incident involving a former US President in Poland (whose State Department interpreter, a Russian specialist, was perceived to be insulting the audience) illustrates the folly of this position, yet it is still widely held. Thus, translation is not particularly respected as a profession in the US — it is frequently perceived as the 'last resort' of otherwise-unemployed liberal arts majors — and the pay is poor (an ironic contrast with the reviving interest in MT in the US).

In Canada, in Europe, and elsewhere around the world, this myopic attitude is not held. Where translation is a fact of life rather than an oddity, it is realized that one translator does not equal another, that any translator's competence is sharply restricted to a few domains (this is especially true of technical areas), that native fluency in a foreign language does not bestow on one the ability to serve as a translator, and that different languages deserve treatment as such. Thus, in college-level and postgraduate schools that teach the theory (translatology) as well as the practice of translation, a translation student is trained in the few areas in which he will be doing translation. True competence is expected only after years of experience beyond that training, and even then only in a few limited domains of expertise.

Of special relevance to MT is the fact that essentially all translations for dissemination (export) are revised by more highly qualified translators who necessarily refer back to the original text when post-editing the translation: unrevised translations are regarded as being inferior in quality, or at least suspect, and for many if not most purposes they are simply not acceptable.

In the multinational firm Siemens, even internal communications that are translated are also post-edited. Such news generally comes as a surprise, if not, indeed, a shock, to most people in the US.

2.3 Translator resistance

Some translators will never be torn from their dictaphones, and will finish their careers penning corrections on drafts transcribed by secretaries. Others have advanced to typing their own translations — some, even, on word processors, where they may also make corrections, though frequently resorting to marking up printed drafts. Many of these others would still not want a computer to translate for them. Thus, there is a large class of translators who could not appropriately be called potential users of MT systems. There seems to be a high correlation of youth with willingness to experiment with MT (or, for that matter, with any new technology).

A second area of resistance to MT is a fear that translation produced by computers may, in the long term, have a detrimental effect on a translator's own style and ability to translate well. Some translators who have already been exposed to working with poor MT systems may also be opposed to the idea because of such unsatisfactory experiences.

2.4 Developers' ignorance

There remain varying levels of misunderstanding among translators and MT system developers about what is involved in translation with machine aids. Translators are becoming increasingly involved in the development of MT systems, but some developers are still not fully aware of the needs and concerns of translators — the end-users of MT systems. It is important for system developers to understand that there are still translators who resist the advent of MT because they fear being replaced by a computer, although this fear does seem to have decreased somewhat in recent years as translators have become more familiar with other new technologies (word processors, on-line terminology databases and so on).

2.5 Lessons for all

It is easy to see that the 'fully automatic high-quality machine translation' standard, imagined by most US scholars past (the ALPAC committee) and present to constitute minimum acceptability, must be radically redefined. Indeed, the most famous MT critic of all, when faced with evidence of significant user satisfaction with seemingly poor translations performed by computers, eventually recanted his strong opposition to MT, admitting that these terms could only be defined by the users, according to their own standards, for each situation (Bar-Hillel, 1971). However, the US research-funding and academic communities have largely failed to notice and learn from this retraction, and for the most part remain convinced of the impossibility of successful translation by computer — ironically, often quoting Bar-Hillel's earlier 'disproofs' as authoritative.

In informed circles, it is now recognized that an MT system does not have to scan books, then print and bind the results of its translation in order to qualify as ‘fully automatic’. ‘High-quality’ does not rule out post-editing, since such proscription would ‘prove’ the infeasibility of high-quality human translation. Academic debates about what defines ‘high-quality’ and ‘fully automatic’ are considered irrelevant by the users of M(A)T systems. What matters to them are two things: whether the systems can produce output of sufficient quality for the intended use (e.g., information acquisition, or revision for dissemination), and whether the operation as a whole is cost-effective or, rarely, justifiable on other grounds, such as speed. (The Georgetown Automatic Translation system, as with other installations of its type, offered the choice of translations available overnight, if produced by machine, vs. delays of several months, if done by human translators.)

There certainly exists some innate translator resistance to MT, and system developers will have to live with this fact; more, they may have to educate their potential customers, who might in ignorance consider a system to be a failure if it is not universally adopted. When we speak of translators, then, as potential users, we are talking about those whose temperament admits such a possibility. Even in this group, there are barriers to overcome, but of course translators deserve proof that MT can improve their work; some systems have admittedly not done so.

It is precisely for these reasons that it is vital for system developers to understand what is involved in the art of translation, as indeed many translators justly view their craft as a creative art. Having gained an insight into the steps involved in translation, system developers can implement MT systems in such a way that they become useful tools for translators, who need no longer view MT as a threat to their livelihood, linguistic integrity, or sanity.

2.6 A more balanced viewpoint

Existing MT systems have already proven that they hold a place in the field of industrial translation. They are powerful tools for translators, who are relieved of some of the more tedious tasks such as repetitious translation, re-typing, or cutting and pasting diagrams involving very little actual translation. In order for MT to gain widespread acceptance, however, it is important for each system to be implemented with the full range of user needs in mind and to include a full range of adaptable and user-friendly tools. An MT system may be based on highly sophisticated linguistic theories, but it will not be acceptable to users unless it is relatively simple to learn and operate, because translators will be unwilling to adapt to the new technology and will return to their individual methods of translation.

That is not to say that translators should not be ready to adapt to MT systems; indeed, they can derive the most benefit from MT by being prepared to learn the new skills required. Translators also have a great deal to offer by becoming involved in the development of systems intended for their use, and by providing continual feedback about their experiences with MT to system developers and implementers.

3. SURVEY

This section presents a cursory overview only; Slocum (1985) offers a more expanded treatment. The tutorial will expand on the general themes presented herein. We divide the MT world into the three most active regions: North America, the Far East, and Western Europe.

3.1 North America

North America, particularly the US, was until recently rather isolated, even insular. The economies were largely domestic and, in the case of Canada and the US, English was sufficient for virtually all government and commercial purposes. With the adoption of French as an official national language, the situation in Canada began to change: the amount of translation required grew dramatically, and translation bureaus soon could not cope adequately with the volume. Thus, a serious interest in MT arose in that country, resulting in part in the TAUM project at the University of Montreal.

Although, through demographic studies, one can foresee potential for the promotion of Spanish as an official national language of the US, that time is well into the future. Only with the recent shift of the US from a domestic toward an international economy has there been a growing perception of translation services as a serious need. Heretofore, US industry, to the extent that it dealt with foreign concerns at all, has generally succeeded in requiring customers to interact in English, and to accept English documentation. The 'translation bureaus' of major firms sometimes consisted of no more than a few bilingual secretaries—and those, of course, having no training as translators. As the US has begun turning to foreign markets—and meeting with stiff competition from international companies which, among other things, are willing and able to deal in customers' native languages — the interest in translation has increased.

This trend, however, is only recent, and quite small. Frequently it is not even perceived until one points out evidence such as the recent establishment of MT projects inside several US computer companies. Even in the last few years, major commercial MT vendors have reduced their domestic staffs and further increased their attention to foreign markets. The only large academic MT project in the US, at the University of Texas, receives its entire support from Siemens, a West German firm; a related, but purely research project at the same institution, receives its funding from Hitachi, a Japanese firm.

It is a paradox, then, that—until two years ago—the major commercial MT firms had all arisen in the US, even though their markets lay elsewhere; perhaps this is a testament to the American entrepreneurial spirit. In any case, this situation too has changed. More than a dozen Japanese companies have announced commercial MT systems, or production-prototypes, since 1984, and the customer base already has numbers in the hundreds. At a time when the need for translation is finally becoming apparent, the US is rapidly losing any edge it may have had in this field.

3.2 The Far East

Unlike the US, Far Eastern countries have not enjoyed the luxury of requiring potential customers to learn their languages. Especially for Japan, whose economy is founded upon the export trade, the ability to provide documentation in the customer's language is of paramount importance. Contrary to popular belief (in the US, at least) about the propensity of the Japanese for learning foreign languages (English), most of them find it an arduous task, and even translation for information acquisition is therefore highly important.

Translation in Japan now constitutes an annual market estimated at around one trillion (1000 000 000 000) yen, with some of the larger industrial firms reporting individual translation budgets exceeding one billion (1000 000 000) yen; perhaps 80% of this comprises technical texts for the export market, of which English accounts for about half (Philippi, 1985). This is a staggering demand, and the prospect of a trillion-yen market, which may double in 2-3 years, is dazzling. It is quite understandable, then, that no small number of Japanese companies are interested in developing, using, and marketing MT systems.

3.3 Western Europe

Europeans have for millenia coped with language problems. Normal commerce requires the ability to communicate in several languages, and companies traditionally use both in-house and outside translators to supply documents to customers. With the establishment of the European Economic Community (EEC), and its subsequent expansion from four to nine official languages, the growth in demand for governmental translation has been staggering. In 1986 the European Parliament spent nearly half its budget for translation and interpretation services. The Council of Ministers employs over 700 translators, and the cost of translation in the Commission exceeds that of Commission-supported research.

In such an environment, it is easy to see how interest in MT can flourish as it does. Accordingly, individual European companies and governments, to say nothing of the EEC itself, are willing to invest large sums in techniques that promise to deal with translation demands more effectively. As a result, there are a number of projects scattered about Europe whose goals are to ameliorate the translation problem in one way or another. Some government-supported projects have been large, and of long standing: the Groupe d'Etude pour la Traduction Automatique (GETA), at Grenoble, and its sister group at the University of the Saar, in Saarbrücken, come to mind. Other groups, some supported by government and some by industry, have been smaller.

Some groups have been effective in the sense of producing systems admitting practical application, and others not; but failures occur in all research and development areas, not just MT. Indeed, lack of success is not always complete: one early group — CETA at Grenoble — failed in its original mission, but learned from its mistakes and, reinstituted as GETA, adopted another approach; this eventually resulted in a prototype now undergoing production implementation in the French national project ESOPE.

4. SYSTEM CLASSIFICATION SCHEMES

In order to appreciate the technologies behind machine translation systems, it is necessary to understand the broad levels of ambition by which they can be categorized, and something about the linguistic techniques that MT systems employ in attacking the translation problem. Each of these, in different ways, has a profound effect on system design. In this section we outline the major design issues.

4.1 Levels of ambition

There are three broad categories of 'computerized translation tools' (the differences hinging on how ambitious the system is intended to be): machine translation (MT), machine-aided translation (MAT), and terminology databanks (TD). Up to now, we have only discussed the first. We shall continue to concentrate on this end of the spectrum, but the reader should remember that there are these other categories. In fact most of our comments throughout this paper, though addressed at MT systems, also reflect on these others.

MT systems are intended to perform translation 'without human intervention'. This restriction does not rule out pre-processing (except for such decisions as marking phrase boundaries and resolving part-of-speech and/or other linguistic ambiguities), or post-editing (since this is normally done for human translations anyway). However, an MT system is solely responsible, without human assistance, for the complete translation process from input of the source text to output of the target text, using special programs, comprehensive dictionaries, and collections of linguistic rules (to the extent that they exist, varying with the MT system). On the scale of computer translation ambition, MT occupies the top position. All of the early work in MT addressed this goal.

MAT systems fall into two subgroups: human-assisted machine translation (HAMT) and machine-assisted human translation (MAHT), reflecting successively lower levels of ambition. In an HAMT system the computer is responsible for producing the translation *per se,* but may interact with a human monitor at many stages along the way, by asking the human to disambiguate a word's part of speech or meaning, or to indicate where to attach a phrase, or to choose a translation for a word or phrase from among several candidates discovered in the system's dictionary. In a MAHT system, on the other hand, the human is responsible for producing the translation *per se* (on-line), but may interact with the system in certain prescribed situations, for example, requesting assistance in searching through a local dictionary-thesaurus, accessing a remote terminology databank, retrieving examples of the use of a word or phrase, or performing word processing functions such as formatting. The existence of a pre-processing stage is unlikely in an MA(H)T system (the system does not need help, rather it is making help available), but post-editing is frequently appropriate.

Terminology databanks (TD) are the least ambitious systems because access frequently is not made during a translation task (the translator may not be working on-line), but usually is performed prior to human translation. Indeed, to the translator, the databank may not be accessible on-line at all, but may be limited to the production of printed subject-area glossaries. A TD offers access to technical terminology, but usually not to common words (the translator already knows these). The chief advantage of a TD is not that it is automated (even with on-line access, words can be looked up almost as quickly in a printed dictionary), but that it is up-to-date: technical terminology is constantly changing, and published dictionaries are essentially obsolete by the time they are available. It is also possible for a TD to contain more entries because it can draw on a larger group of active contributors: its users.

4.2 Linguistic techniques

We shall characterize the methodologies employed for MT along two axes: depth of analysis, and the primary source of knowledge. Regarding the depth of analysis, there are three classic categories (direct, transfer, and interlingua). Regarding the primary source of knowledge, there are four (lexical, syntactic, semantic, and cognitive). We shall consider these two axes in turn.

4.2.1 Depth of analysis

Superficially, translation may be regarded as analysing an input (text, sentence, or whatever) in the source language (SL), and synthesizing an equivalent output in the target language (TL). In designing an MT system, one must decide how ‘deep’, or comprehensive, the analysis is to be. This will determine where one starts when it comes time to perform synthesis, and hence the design of the entire synthesis module.

At the shallowest level, one can perform ‘direct translation’, which is characterized by the fact that analysis of the SL is restricted to the minimum work necessary to produce a translation in a single, specific TL. Assuming that such a shallow analysis can be sufficient for high-quality translation (and there are arguments concerning whether or not this is true), it remains the case that another translation of the same input, into a second TL, requires a complete re-analysis of that input. This is considered to be inefficient, from both the operational standpoint (performing a new analysis for the new TL) and, perhaps more importantly, the standpoint of the development project itself (since, in theory, one can use nothing from one linguistic rule base when building another, each being geared toward a single, specific language pair). Consider, for example, that a direct MT system contains in the SL dictionary details of the behaviour of the TL: for translation into a new TL, one must construct an entirely new dictionary. For translation in all directions among N languages, one would have to construct $N \times (N-1)$ entirely different MT systems: 72 of them, for the 9 EEC languages.

At the other extreme, one can opt to analyse an input into a deep, language-independent representation of meaning usually called an interlingua. From this meaning structure, one can in principle translate into multiple languages merely by synthesizing multiple outputs — without ever re-analysing the input. Thus, for translation in all directions among N languages, one need construct only N analysers plus N synthesizers. To use our dictionary example, only N dictionaries (one per language) are required. This would seem to maximize operational and development-project efficiency. One major problem with this is that linguists are not yet able to specify an interlingua: the necessary theories have not yet been devised (indeed, that one can possibly exist is mere conjecture). Another problem is that, the world being an imperfect place, there will always be some sentences that cannot be analysed — either because the sentence is truly ungrammatical or, more likely, because the machine's grammar is incomplete. In a system relying on an interlingua, a failure to analyse fully almost necessarily entails a failure to synthesize, and no translation is produced — something human users do not tolerate well.

A compromise position involves an intermediate stage called transfer. An input is analysed into a structure deeper than that required for direct translation, but not (quite) language-independent like an interlingua. Then a transfer step transforms that structure (peculiar to the SL) into an equivalent structure peculiar to the TL, and the synthesizer uses the result to produce the output translation. In our hypothetical N-language environment, this requires N analysers, and N synthesizers, plus $N \times (N-1)$ transfer modules — which seems the worst of all solutions. (One needs N monolingual dictionaries, and at least $N \times (N-1)/2$ bilingual dictionaries.) However, it is generally agreed that the transfer modules and bilingual dictionary entries can be kept sufficiently small that the research and development efficiency is acceptable. This approach also has the great advantage that it has been shown to work well in practice — better than the direct approach, which was generally abandoned years ago — whereas the interlingua approach has thus far failed to produce a workable system, and clearly will not do so in the near future. Today, essentially all serious MT system development efforts employ the transfer approach, while the few extant long-term MT research efforts tend to concern themselves with the interlingua approach.

4.2.2 Primary source of knowledge

Another design decision faced by an MT system developer is the locus of the primary source of knowledge used to guide the translation. One could locate most knowledge about language behaviour in the lexicon — as, for example, a direct translation system does. This tends to result in maximally specific rules (hence, presumably, the highest accuracy). However, it vastly increases the size of the rule base unless one is extremely clever about organizing a lexical taxonomy supporting feature inheritance — something not yet demonstrated. Worse, in practice this approach has led to many kinds of unforeseen, destructive rule interactions.

An alternative locus of the primary source of knowledge is in syntax rules. Almost all of the most promising MT systems today rely primarily on this form of knowledge. However, it is difficult if not ultimately impossible to control a purely syntactic knowledge base, and all these systems use other sources of knowledge to reduce what would otherwise be an explosive growth of interpretations as the rule base and sentence length grow. Syntax rules also suffer from problems of fragility: for example, there is little or nothing one can do to recover a full-sentence analysis of an input sentence that is structurally unsound by the standards of the syntax rules.

A third potential locus is in the semantic rules. It is claimed that these can be somewhat less fragile than syntax rules in the presence of errors, though in practice this has not been remarkably obvious. However, no semantic-rule formalism has achieved widespread acceptability, for the reason that all proposals are known to be faulty in too many respects. It is also the case that currently-proposed semantic rules tend to be rather large and time-consuming to construct, individually, with the result that the manpower investment required to produce a practical application is prohibitive.

With respect to the fourth potential locus of primary knowledge, cognitive rules (including, e.g., common-sense knowledge of the world), the same arguments apply, only to a greater extreme. They are potentially the most robust, if proponents are to be believed, but they are easily the most debated, complex, and costly form of knowledge yet proposed. It is hard to imagine constructing any large application by encoding their requisite cognitive knowledge base manually. (Consider the world knowledge subsumed by a 100 000-page telephone switching-system manual.) Rather, it is more likely that machine learning of such structures will be required before they can be used outside of 'toy' problem domains. MT systems are generally intended to be applied to hundreds of thousands, perhaps millions, of pages of text, whereas NLP systems based on cognitive rules have been 'applied' to texts measured in tens of words (that is, where any form of comprehension of the entire text was attempted: we discount what amount to augmented keyword searches posing as models of cognitive processing).

5. LOOKING AT TRANSLATION

Before one can meaningfully critique a translation, or a translation process, whether performed by man or machine, one must determine the use to which the result is being put. When considering, further, whether translation by machine is necessary, or even desirable, one must recognize the strengths and weaknesses of computers vs. humans as translators, vis-à-vis the type of text.

5.1 The purposes of translation

Consider two extreme purposes of translation: information acquisition vs. information dissemination. The classic example of the former is intelligence-gathering: with masses of data to sift through, there is no time, money, or incentive to translate carefully every document by normal (i.e., human) means. Scientists more generally are faced with this dilemma: already more must be read than can be read in the time available. Having to labour through texts written in foreign languages—when the probability is low that any given text is of real interest—many not be worth the effort. In the recent past, English has been the lingua franca of science, but it is becoming less and less dominant for a variety of reasons, including the rise of nationalism and the spread of technology around the world. As a result, scientists who rely on English are having greater difficulty keeping up with work in their fields. If a very rapid and inexpensive means of translation were available, then — for texts within the reader's areas of expertise — even a low-quality translation might be sufficient for information acquisition. At worst, the reader could determine whether a more careful (and more expensive) translation effort might be justified. More likely, his understanding of the text would be good enough to preclude the need for a more careful translation. The older MT systems were generally intended for this application. Without them most documents are never translated because, in economic terms, producing careful translations that will only be cursorily scanned (even assuming they are not outdated by the time they are available) is a waste of resources.

The classic example of the latter purpose of translation is technology export: an industry in one country that desires to sell its products in another country must usually provide documentation in the purchaser's native language, or at least a designated second language. In the past, US companies, for example, have escaped this responsibility by requiring that the purchasers learn English; other exporters (German, for example) have never had this kind of luxury. With the increase of nationalism, it is likely that monolingual documentation will be less acceptable; for this and other reasons, translation is becoming increasingly common as more companies look to foreign markets. More to the point, texts for information dissemination (export) must be translated with a great deal of care: the translation must be 'right' as well as clear. Qualified human technical translators are hard to find, expensive, and slow (translating somewhere around 4-8 pages/day, on the average). The information dissemination application is most responsible for the renewed interest in MT.

5.2 Literary vs. technical translation

Natural language texts range in 'linguistic complexity' from technical documentation through edited abstracts and scientific reports to newspaper articles, literary works, and political-sales material.

Literary and technical translations, residing as they do near opposite extremes of this range, differ with respect to both process and output. In literary translation, emphasis is placed on style, perhaps at the expense of absolute fidelity to content (most especially for poetry). In technical translation, emphasis is properly placed on fidelity — especially with respect to technical terminology — even at the expense of style. Highly polished technical translations, especially of manuals, are often not considered worth the investment required to produce them. (Indeed, highly-polished original texts are rare!) Technical accuracy is a critical consideration.

As far as MT is concerned, technical is less complex than literary translation because it is characterized by relatively less syntactic and semantic variety and more denotative as opposed to connotative content. Paradoxically, the order of complexity for human translators is essentially reversed for reasons of vocabulary: the acquisition, maintenance, and consistent use of valid technical terminology is an enormous problem. No qualified human translator has much difficulty with straightforward syntax or normal idiomatic usage, but the prevalence and volatility of technical terms and jargon poses a considerable problem. MT systems may lack good style, but they excel at terminological accuracy and speed: they are best suited for technical translation.

As it happens, there is little or no need for literary translation by machine: there is no great shortage of human translators capable of handling the load. By contrast, in many technical fields there are far too few qualified human translators, and it is obvious that the problem will never be alleviated by such measures as greater financial incentives, however laudable that may be.

The demand for technical translation is staggering in sheer volume. A single set of operation and maintenance manuals for, say, a modern telephone switching system can exceed the size of all of the works in classical Greek literature. Also, as the manuals are revised, they may require retranslation several times. The only hope for a solution to the technical translation problem lies with increased human productivity through the full range of computer technology: full-scale MT, less ambitious MAT, on-line terminology data banks, and word processing.

5.3 Advantages in technical MT

There are several reasons why computers are already effective at translating technical texts. One of theses concerns vocabulary: technical texts (hence, MT system dictionaries) tend to concentrate on one subject-area at a time, wherein the terminology (lexical semantics) is relatively consistent, and where the vocabulary is relatively unambiguous, even though it may be quite large. (This is not to say that lexical problems disappear!) Another advantage is that there is typically little problematic anaphora, and little or no ‘discourse structure’ as usually defined; thus MT systems can usually get away with ignoring these hard problems. Third, in accordance with current practice for high-quality human translation, revision is to be expected. That is, there is no *a priori* reason why machine translations must be ‘perfect’ when human translations are not expected to be so: it is sufficient that they be acceptable to the humans who revise them, and that the translation process proves cost-effective overall (including revision).

5.4 Problems in technical MT

Notwithstanding the advantages of MT for technical texts, there are definite problems to be confronted. First of all, the volume of such material is staggering: potentially hundreds of millions of pages per year. Even ignoring all cost-effectiveness considerations, the existence of this much candidate material demands a serious concern for efficiency in the implementation. Second, the emphasis in MT is changing from information acquisition to information dissemination. The demand is not so much for loosely approximate translations from which someone knowledgeable about the subject can infer the import of the text (perhaps with a view toward determining whether a human translation is desired); rather, the real demand is for high-quality translations of, e.g., operating and maintenance manuals — for instructing someone not necessarily knowledgeable about the vendor's equipment in precisely what must (and must not) be done, in any given situation. Fidelity, therefore, is essential.

In addition to the problems of size and fidelity, there are problems regarding the text itself: the format and writing style. For example, it is not unusual to be confronted with a text which has been 'typeset' by a computer, but for which the typesetting commands are no longer available. This can be true even when the text was originally produced, or later transcribed, in machine-readable form. The format may include charts, diagrams, multi-column tables, section headings, paragraphs, etc. Misspellings, typographical errors, and grammatical errors can and do appear. Technical texts are notable for their frequency of 'unusual' syntax such as isolated phrases and sentence fragments, a high incidence of acronyms and formulae, plus a plethora of parenthetical expressions. The discourse structure, if it can be argued to exist, may be decidedly unusual — as exemplified by a flowchart. Unknown words will appear in the text. Sentences can be long and complicated, notwithstanding the earlier statement about reduced complexity: technically-oriented individuals in all cultures seem to be renowned for abusing their natural languages. The successful MT system, then, will address these problems as well as those more commonly anticipated.

6. BUILDING A PRACTICAL SYSTEM

Constructing a practical MT system is extraordinarily difficult. Many questions arise to which there are few known (or at least agreed-upon) answers. Certain of these questions, being of academic interest, are discussed in considerable detail in the open literature. Matters of the parsing algorithm, the linguistic formalism, the translation strategy, etc., will thus be ignored here. One critical area, however, being of mere practical interest, has received very little scholarly attention: the user interface. By 'user' we refer, variously, to the linguist who is developing the grammar rules; to the lexicographer-terminographer who is developing the dictionary entries; and to the translator who is of course the true end-user. In our discussion below, we use the term 'linguist' to refer to anyone engaging in a strictly system-development role (e.g., grammarian, lexicographer); a 'translator' is anyone engaging in a pre- or post-editing or terminology update role.

No natural language processing (NLP) system is likely to be successful in isolation. An environment of support tools is necessary for ultimate acceptance on the part of prospective users. The following 'linguist support tools', we think, constitute a minimum workable environment for both development and use of NLP systems generally: a DBMS for handling lexical entries; validation programs that verify the admissability of all linguistic rules (grammar, lexicons, transformations, etc.) using a set of formal specifications; dictionary programs that search through large numbers of proposed new lexical entries (words, in all relevant languages) to determine which entries are actually new, and which appear to replicate existing entries; defaulting programs that 'code' new lexical entries in the NLP system's chosen formalism automatically, given only the root forms of the words and their categories, using empirically-determined best guesses based on the available dictionary database entries plus whatever orthographic information is available in the root forms; and benchmark programs to test the integrity of the NLP system after modifications (Slocum, 1982). A DBMS for handling grammar rules is also a good idea.

For MT applications, one must add 'translator support tools': a collection of text-processing programs that (semi-)automatically mark and extract translatable segments of text from large documents, and which automatically insert translations produced by the MT system back into the original document, preserving all formatting conventions such as tables of contents, section headings, paragraphs, multi-column tables, flowcharts, figure labels, and the like; a powerful on-line editing program with special capabilities (such as single-keystroke commands to look up words in on-line dictionaries) in addition to the normal editing commands (almost all of which should be invokable with a single keystroke); and also, perhaps, (access to) a 'term databank', i.e., an on-line database of technical terms used in the subject area(s) to be covered by the MT system.

The first section below deals with support tools for the linguist, which aid in the development and maintenance of any NLP system. Next we deal with tools for translators, which aid in the document preparation, translation, and editing tasks.

6.1 Tools for linguistics

There are several issues that must be addressed in the selection and implementation of natural language processors and their accompanying support environments. The computing world is slowly but surely coming to recognize the importance of what is called 'the programming environment' in research and development circles, or 'integrated software' in application circles. This is opposed to an earlier focus on, e.g., the hardware or programming language *per se*. We shall therefore open out discussion with a consideration of some man-machine interface issues relating to the dictionary-grammar production cycle. In particular, we take the position that a rule interpreter cannot be separated entirely from its system development and maintenance environment; rather, it must fit naturally into its environment so that overall productivity may be maximized.

As our canonical linguistic rule interpreter, we designate the parser, though of course there are other things one does with (other kinds of) linguistic rules. By the term 'parser', we refer to a program that interprets a 'sentence' according to a distinguished set of dictionary and grammar rules; that is, a parser does not itself incorporate the linguistic rules by which sentences are interpreted. Although adhering to good principles of modular programming may cause a parser to be somewhat divorced from its operational environment, the separation cannot be rigid. In particular, one should not consider a parser with total disregard for its operating environment. One must consider the overall NLP system, including the facilities that support the development and maintenance of the linguistic rules (dictionaries and grammars) that the parser relies on, the format and structure of the rules as seen by the lexicographer and linguist, the testing facilities that allow one to evaluate the NLP system, and the relationship of the parser to that system.

6.1.1 Rule development and maintenance

In any large software system, the problem of producing and maintaining the data sets on which the programs operate becomes important, if not critical. First of all, when there is a large volume of such material the data-entry process itself can consume a significant amount of time; second, the task of ensuring data integrity becomes an even larger time sink. We shall expand briefly on these two problems.

There are two problems associated with data entry: creating the data in the first place, and getting it entered in machine-readable form. In most applications of database management systems, the creation of data is relatively straightforward: such data items as personal name, age, job title, identification number(s), salary, etc., serve as examples to illustrate the point that the data items usually pre-exist; thus data entry becomes relatively more important. In an NLP application, however, creating the original data is the major bottleneck: one must decide, for each of thousands of . words, many details of behaviour in a complicated linguistic environment. Certainly these data may be said to 'pre-exist' (in the language), but a real problem arises when humans attempt to identify them. In general, the more sophisticated the NLP system, the more of these details there are. Data entry, relatively speaking, becomes a small or insignificant issue—although it remains a significant issue in absolute terms.

In the dictionary realm, an initial solution to this problem is a 'lexical default' program that accepts minimal information (the root form of the word, and its category) and automatically encodes the features and values that specify the details of linguistic behaviour. This can be accomplished by a combination of morphological analysis of the root form of the input word, search of the existing dictionary database for 'similar' entries, and statistically-justified guessing. This amounts to automated dictionary-entry coding by analogy. Defaulted lexical entries can be created in machine-readable form to begin with, and made available for human review-revision using standard or special-purpose on-line editing facilities. This can reduce greatly both coding time and coding errors. It does not seem that any similar solutions to this problem yet exist for grammar rules, though one can imagine a system researching for grammar rules analogous to the one desired.

A potentially harder problem, however, is the maintenance of data integrity. Humans will make mental errors in creating lexical entries, and will aggravate these by making typographical errors during data entry. Even assuming a lexical default program (which, of course, does not make such 'syntactic' mistakes), the process of human revision of the defaulted entries may introduce errors. The solution to this problem would be to include a validation program that, working from a formal specification of what is legal in linguistic rules (grammatical or lexical), identifies any errors of rule format and/or syntax within each submitted entry. The formal specification could be organized by language, by lexical category within language, and perhaps by subcategorization feature.

There is also the problem of maintaining an existing database. In any NLP system, there will be a need for changing linguistic rules in the light of experience; in a system that serves as a vehicle for research in NLP, this problem is magnified by the occasional need for large-scale changes to accommodate new system features, or even theories of language. One approach to this problem is to incorporate a general database management system (DBMS) along with a group of interface routines that transform, upon entry, rules from a format optimized for human use into a format more suitable for storage by the DBMS, and which reverse the transformation when retrieving the rules.

Finally, there is the problem of trying all these modules together with a powerful, high-level user interface that optimizes the task of rule acquisition and maintenance. Programs intended to solve this problem should use the database interface, the default program (in the case of lexical entries), the formal specifications governing linguistic rules, and the rule validator, to facilitate the process of rule acquisition and maintenance. Such programs must support the usual requirements for adding, modifying, and deleting all types of rules.

6.1.2 Rule structure and format

The market for computational linguistics is growing more rapidly than the number of trained personnel who can best fill the positions. Even if this were not true, other factors would argue for the NLP system grammars and lexicons being maintained in a format optimized for use by linguists and lexicographers, rather than by programmers *per se.* The issue of overall efficiency in research, development and application precludes interest in machine efficiency alone. NLP is an exceedingly difficult problem whose optimal solution is not yet well understood: empirical results can and will dictate that linguistic procedures proliferate and be changed. For this to be effected by linguists and lexicographers who are not sophisticated in the computer arts, the rule bases must be expressed in a formalism with which they are familiar, or at least which they can easily understand. This tends to eliminate, e.g., LISP code from consideration.

Similarly, the software component (parser) should impose no significant constraint on how the linguistic component represents interpretations of sentences. The most common representation formalism in modern linguistics in the US, for example, is constituent-structure trees; in related disciplines, other formalisms are preferred. In order to allow freedom of choice, a few specialized routines could be written for each desired representation; the parser should interface with these modules 'at the back end' in a well-defined manner. Thus the linguists could change their representation formalism at will.

6.1.3 Rule base testing and evaluation

Linguistic rules, like programs, require extensive debugging and tuning. Since the phenomena that one is trying to account for in an NLP system are generally more complex and open-ended than those encountered in other forms of programming, it can be argued that the need for debugging-tuning tools is even more acute. It is strange, then, that the field has witnessed so little discussion of such issues. Perhaps this is due to the fact that most NLP systems deal with very small subsets of natural language, so that all of the behavioural characteristics of the system can be accommodated in the computational linguist's head. At any rate, when NLP systems grow large, human memory fails; such tools then become indispensable, else the system falls apart.

Certainly, the minimal information required for debugging is a trace of the applications of all rules (lexical and grammatical). Both the input and output of each rule should be noted, including reasons for failure. In any large system, where the number of rule applications grows rapidly with the increase in rule-set size, the tracing behaviour should be selective — conditioned on, e.g., information type (lexical, syntactic, semantic), and/or such information content as category (noun phrase, clause, etc.). Otherwise, the flood of information will be more than the linguist can effectively deal with.

For tuning purposes, the needs are greater. In addition to recording the application history of all rules, special analysis programs must be written to summarize and present the data thus gathered. Data points should include such things as, for each rule, the number of applications attempted, the number of successes or failures that result in a local sense (conditioned on subcategorized features, including semantic tests), and the number of times a construction actually appears somewhere in a global (complete sentence) analysis.

6.2 Tools for translators

This section discusses the ways in which end-users interact with MT systems, and the need for MT system developers to be aware of user requirements. A collection of adaptable techniques to facilitate the operation of an MT system at each stage of interaction is proposed.

A user interacts with an MT system in four different stages:

(1) text preparation,
(2) dictionary update,
(3) execution of translation run, and
(4) post-editing MT output.

Each of these four stages requires special tools so that the relevant task may be accomplished as easily and effectively as possible.

6.2.1 Text preparation

Whenever a text is to be translated on an MT system, it is necessary for the text to be available in machine-readable form, but a text submitted for translation may be supplied as a typewritten or even hand-written document. This first step of inputting a text can prove to be more than a trivial problem: MT system developers need to be aware of this possibility, and to make available text entry facilities that are as powerful and as user-friendly as possible.

Once the text is available in machine-readable form, it must be prepared for translation. Many technical documents contain diagrams in which blocks of text do not run strictly from left to right; these must be marked and extracted as separate translation units, then re-formatted when the translated document is produced. An MT system should recognize, translate, and re-format all diagrams: if too much time is spent on manual pre-processing, the cost of translation begins to rise, the advantages of speed and cost-effectiveness when using MT begin to decline, and user-acceptability decreases.

6.2.2 Dictionary update

Once the text is available to the MT system, it is usually necessary for any new words that do not yet appear in the system's dictionary to be added. User acceptance of the dictionary update procedure tends to involve two factors:

(a) the kinds of information that a user is required to supply, and
(b) the method used to enter that information into the dictionary.

Even during the linguistic development phase, consideration should be given to the kinds of information that the user will be asked to code in dictionary entries. If the coding process becomes too complex, the user will be frustrated and the number of incorrectly-coded dictionary entries is likely to increase. It may be necessary for MT system developers to avoid a powerful but complex system, considering the ability of the user to code correctly the dictionary entries. It is important that the user be able to understand the effects of lexical coding decisions on the resulting translation: not only does this improve the quality of dictionary entries, but it also enables the user to gain more control over the dictionary and, consequently, the quality of the translation. Experience has shown that acceptance by users is enhanced when they have some control over the output of the MT system (Piggott, 1982).

6.2.3 Running the translation

Once the source text has been made available to the MT system and the dictionary update is complete, translation can begin. If users are to manage the execution of these tasks, there must be user-friendly tools, including a mnemonic means to request execution of them, to check on their status, and to be informed upon the completion of a translation. In order to gain the most benefit from an MT system, the system must accept requests for translations to be run in batch mode (e.g. overnight).

In general, the aims of developers should be to produce an MT system in which the task of scheduling translations is very simple, and in which translation jobs can be executed efficiently in order to make the best use of system resources.

6.2.4 Post-editing

An editor may need to revise a translation and, since user acceptance of MT is vital to its ultimate success, it is especially important that the system tools developed for post-editing be user-friendly and flexible. Post-editing tools should be designed so that a revisor can make changes to the draft translation with a minimum number of keystrokes: this will enable him to work rapidly, and will minimize frustration.

The translator should be able to select his preferred method for editing. Some prefer to write changes on a hardcopy printout of the translation and to input these changes into the computer later; others prefer working on-line, but with a hardcopy of the source document available for reference; still others prefer working entirely on-line, with alternating sentences of source and target text, or with two windows containing source and target text. Some prefer to re-format a text before editing, while others would re-format afterwards; this may depend on the type of text that is being post-edited, and whether it contains flowing text or a large number of diagrams and tables. An ideal MT system would allow users to edit using the method they find most comfortable.

Although there are many powerful word processing systems available, most of these were designed with text input in mind rather than post-editing. The particular changes that a post-editor needs to make to a document depend on the accuracy of the draft translation, but there are some general changes that appear to be common to all MT systems. A post-editor tends to work in units of sentences and, therefore, sentence-oriented functions that allow the cursor to be moved to the beginning or end of a sentence, as well as of a line, are necessary.

One of the most frequent post-editing changes made to translations is modifying word order. A screen editor for an MT system should, therefore, also include a full range of functions for moving-deleting words with single keystrokes. An ideal MT system would also include simple ways for entire phrases to be marked and moved from one point in the sentence to another.

Piggott (1982) noted that MT systems tend not to be totally accurate in capitalization. Even if MT systems were to commit no such errors, there would still be cases where a post-editor would break a long sentence into two shorter sentences, or combine two short sentences. Changing the case of a single letter should be implemented as a single-keystroke function; similarly for an entire word.

There will be instances in which a particular term is translated incorrectly throughout a text. The editor should have available two separate global replacement functions: an 'all-at-once' version, for when the editor is confident of the change; and an interactive 'query and replace' version, for substitution on a case-by-case basis (after confirmation).

The user should bear in mind that as much as possible of the original draft translation should be retained; otherwise, the purpose of MT is lost. There are instances, however, when the post-editor must rewrite major parts, or even all, of a sentence. The screen editor should also be able to cope with these cases in a user-friendly way, allowing the editor to overwrite, or alternately to insert new text into, the body of a sentence; ideally, the post-editor would be able to switch back and forth from overwrite-mode to insert-mode with one keystroke.

To allow even greater flexibility, an editing environment could also supply a set of user-programmable keys. These would allow individuals to program the strings of commands used most frequently. For example, if an MT system tended to make errors by inserting extra definite articles in the English target text, a post-editor could program one key to search through the document for each instance of 'the' and a second key to delete it automatically if it was superfluous.

While there is no doubt that new MT system users must learn new skills and accept different kinds of translation problems from those they have been accustomed to, their freedom to select a preferred method for editing a draft translation would enhance the acceptance of MT systems. In addition, the range of screen-editing functions outlined above would decrease the number of keystrokes required to make the necessary changes in a document. This in turn, would speed up the post-editing phase and help to minimize the final cost of working with MT by simplifying the interaction between the user and the system.

7. STATE-OF-THE-ART SUMMARY

Machine translation has long since reached the point of commercial viability: the 15-20 year longevity of some commercial vendors attests to this, as much as anything can. The feverish level of recent activity in Japan, as well as the expanding markets in the rest of the world, indicates that MT is in a growth phase. In this final section we critique the state of the art in terms of the current lines of research, areas receiving too little attention, some of the (still) controversial questions that plague the field, and some problems whose solutions, if available, would result in dramatic improvements.

7.1 Lines of research

A high level of interest has always existed for interlingual (AI) representations of meaning, and research continues along this front. There are two aspects of interlingua: structural, and lexical. So far, objectively discernible progress has been meagre, and research has been plagued by some lack of awareness of the full import of the distinction above. Poor research on interlinguas, of which there is unfortunately too much (frequently conducted by monolingual language speakers, speakers of only closely related languages, or at best linguistically naive computer scientists) indicates nothing, of course. Also unfortunate are proposals to employ a natural language as an interlingua; competent translators must derive considerable amusement from such suggestions. The best research indicates clearly that far more work will be required before this question yields an answer.

Another line of active research in MT concerns grammatical theories, and corresponding representations of grammatical knowledge. Variations on phrase-structure grammar, though criticized as weak, have been popular in the past owing to their computational tractability and relative ease of maintenance; but while useful for analysis and synthesis, there is nothing in the theory that speaks to the problem of transformation—as is required for transfer, or conversion into a semantic representation (e.g., interlingua). Tree-to-tree transducers have been equally popular because of their greater power and straightforward transformational application, but have tended to suffer computationally as well as from maintenance problems. (Since anything is possible, it is hard to know where to start fixing a problem, or extending coverage.) Recently, interest has grown in variants of functional unification grammar — especially within the EUROTRA project. Active investigation of this formalism is increasing, but it is not clear how the formalism could be used for, e.g., transfer.

7.2 Research gaps

Typically embodied in functional unification grammar theory, as well as certain other theories of language, is the notion of lexicon grammar: rules of grammatical behaviour are to be attached to lexical entries. This is all fine and well, provided that one is willing to conduct the large-scale effort necessary to identify those behaviours. Unfortunately, far too few have shown such an interest, and this represents a major gap. Such work is detailed, perhaps (to some) to the point of monotony, but it must take place before one can responsibly claim a breakthrough in grammatical formalisms, or know whether a workable approach has been found.

There have been too few (almost no) systematic contrastive studies of language; rather, it is too often the case that language differences are resolved in some *ad hoc* manner as they are encountered during the course of an MT application. The early EUROTRA work is a start in this direction, but very little of it has been published, and in any case this represents but a drop in the bucket for the nine EEC languages, to say nothing of the language families that are entirely neglected. This is another area where a large-scale, detailed effort is the only answer.

Terminology represents another research gap in the sense that too few MT groups are concerned with it. Precious few MT projects have even considered employing existing term banks, much less investigated the extent to which they might be useful. A frequently automatic assumption is that there is little if anything of value to be gained from the effort because of the naiveté of the linguistic knowledge contained. However, blindly acting on this assumption allows one to lose other perspectives — namely, that good term banks are not assembled casually: considerable attention is paid to several principles of their organization (hierarchical subject-area coding, vendor- or product-specific terminology, etc.), and MT workers would benefit from attending to them.

7.3 Controversial questions

One of the most obvious of the controversial questions surrounding MT is the matter of pre-and post-editing requirements. In the US, where ignorance of good translation practice is the rule, the very notion of post-editing a translation is anathema. Human translators are imagined to work without benefit of editing, and thus MT systems are imagined to be acceptable only in an environment devoid of editing. Even if it were not the case that human translations are typically edited, this argument would not hold. What really matters, of course, is whether MT, including any post-editing, is cost effective or justifiable on other grounds, such as speed. Undeniably it is, which accounts for the substantial and growing interest in the technology. MT research efforts are concentrated on improving the speeds of MT systems, while at the same time reducing the amount of post-editing required, in the interest of further improving the cost-benefit ratio.

A related argument concerns pre-editing. Some form of pre-editing texts to be translated by machine is typical, even if confined to marking sentence boundaries or simply excising materials not to be translated (formulas, etc.). In the case of the Japanese commercial systems, such editing is far more involved, consisting of at least manually resolving lexical and structural ambiguities, if not indeed rewriting the text. No one in MT would like to claim that rewriting a text is desirable, but — again — the relevant question concerns the cost-benefit ratio.

On less obviously practical grounds, the notion of an interlingua is controversial. First of all, there is the question of whether one can exist. Certainly none has been identified. Some Japanese systems are claimed to employ an interlingua. However, when one reads their descriptions including, most revealingly, the example-based arguments, it is clear that the Japanese ‘interlinguas’ are, so far, *ad hoc* rather than theoretical solutions. In some cases this is admitted; in others, the writers seem unaware of the real issues.

The second question related to the use of an interlingua concerns its desirability compared with the alternative (the transfer approach, since history has dealt with questions regarding the direct approach). The lessons of CETA should not be ignored: if an interlingual representation is actually achieved, one loses (by definition) all source-language clues about how to render the translation stylistically — critical information, as any human translator will attest. Finally, there is the matter of practicality: do any benefits of the interlingua approach actually pay for the cost incurred? This question cannot be answered definitively until a true interlingua-based system is available; meanwhile, there is no compelling reason to assume the affirmative.

A related question concerns the necessity for AI-ish techniques (e.g., world models). Whatever the arguments—and there are very good ones—for 'full understanding' being prerequisite to 'high-quality' translation, it remains the case that translators to a considerable extent work by lexical substitution and syntactic rearrangement. That is, translation is not by any means a simple matter of understanding the source text, then reproducing it in the target language — even though many translators (and virtually every layman) believe this is so. On the one hand, there is the serious question of whether, in for example the case of an article on front-line research in semiconductor switching theory, or particle physics, a translator really does fully comprehend the content of the article he is translating. One would suspect not. Johnson (1983) makes a point of claiming that he has produced translations, judged good by informed peers, in technical areas where his expertise is deficient, and his understanding, incomplete. On the other hand, it is also true that translation schools expend considerable effort teaching techniques for low-level lexical and syntactic manipulation — a curious fact to contrast with the usual AI 'full comprehension' claim. In any event, every qualified translator will agree that there is much more to translation than simple analysis + synthesis (an almost *prima facie* proof of the necessity for transfer). World models are not the solution to all translation problems, as some AI proponents would have one believe; the question is whether, and to what extent, they constituted a necessary part of the solution.

Finally, one must consider that even semi-objective MT evaluation and system comparison methods are at best suspect, and at worst nonexistent. Evaluation of translation is, to be sure, inherently subjective. However, there has been virtually no effort on the part of MT workers, commercial or academic, to standardize evaluation techniques, and meaningful comparisons have not been performed. Vendors naturally advance performance figures, including translation accuracy, showing their systems to be the best (at least in some context), but such self-serving claims are not credible. The matter of what might constitute reasonable evaluation criteria, in the context of necessarily subjective judgements about translation quality, is thus very much open to question.

7.4 Unsolved problems

One of the major unsolved problems in MT (indeed, in AI applications more generally) concerns system construction efficiency. Highly-trained experts spend much of their time trying to build these large, complex systems, while their efforts almost certainly could be simplified considerably by the existence and use of good system-building tools. Good general software development tools are just beginning to appear, as it is recognized that human costs — no longer machine costs — are the major contributing factor to system-development expense. Certainly such tools would benefit MT system developers; but equivalent tools for linguists, who are charged with developing large, complex systems of linguistic rules, do not yet exist.

A related problem concerns scale-up bottlenecks. Aside from the system maintenance aspects, discussed above, there are performance aspects. Small systems may appear to function perfectly adequately, in their limited environments; but unless there is some means of predicting behaviour in a scaled-up implementation, trying to develop a fast production version is like shooting in the dark. MT workers could make good use of techniques for predicting future system performance, but none has yet been identified for systems of linguistic rules.

For those involved in applications of current technology, there are linguistic problems related to case roles and semantic markers; specifically, which set to use. There is no standard repertoire. Ultimately the identification of such details becomes a research question, but there is a significant problem relating to the identification of a more-or-less standard collection for interim use, and each project mounts its own effort to select a set.

Finally, there is the problem of world knowledge representation-use. Not only are there no standard representation schemes, but there is no consensus concerning their practical application. Ultimately this boils down to identifying the right questions to ask, before knowing how to search for the answers, so the solution is not soon forthcoming. However, if decent world knowledge representation-use schemes were in hand, there would remain the problem confronting MT developers now: how to integrate such schemes into an MT system, where linguistic knowledge is as important as so-called world knowledge, and is of a very different kind.

REFERENCES

Bar-Hillel, Y. (1971) Some reflections on the present outlook for high-quality machine translation. In W. P. Lehmann and R. Stachowitz (eds.), *Feasibility Study on Fully Automatic High Quality Translation, Final Technical Report RADC-TR-71-295,* Linguistics Research Centre, University of Texas at Austin.

Johnson, R. L. (1983) Parsing — an MT perspective. In K. Sparck. Jones and Y. Wilks (eds.), *Automatic Natural Language Parsing.* Ellis Horwood Ltd., Chichester, West Sussex, England.

Philippi, D. L. (1985) *Japan Intelligence,* Vol. 1, Part 4, Donald L. Philippi, San Francisco.

Piggott, I. M. (1982) The importance of feedback from translators in the development of high-quality machine translation. In V. Lawson (ed.), *Practical Experience of Machine Translation,* North-Holland, Amsterdam, pp. 61-73.

Reifler, E. (1958) The Machine Translation Project at the University of Washington, Seattle. *Proceedings of the 8th International Congress of Linguistics, Oslo.*

Reynolds, A. C. (1954) The Conference on Mechanical Translation. *Machine Translation* **1** (3), 47-55.

Slocum, J. (1982) The LRC machine translation system: an application of state-of-the-art text and natural language processing techniques. *9th ICCL (COLING 82), Prague, Czechoslovakia.*

Slocum, J. (1985) A survey of machine translation: its history, current status, and future prospects. *Computational Linguistics* **11** (1), 1-17.