Machine Translation

Sergei Nirenburg and Yorick Wilks

Abstract. This paper gives an intellectual overview of the field of machine translation of natural languages. At 50, this field is one of the oldest nonnumerical applications of computers. Over the years, MT has been a focus of investigations by linguists, psychologists, philosophers, computer scientists and engineers. It will not be an exaggeration to state that early work on MT contributed very significantly to the development of such fields as computational linguistics, artificial intelligence and application-oriented natural language processing. *Advances in Computers* followed the development of MT closely; a seminal article on MT by Yehoshua Bar Hillel was published in its very first issue. This paper surveys the history of MT research and development, briefly describes the crucial issues in MT, highlights some of the latest applications of MT and assesses its current status.

1. Introduction

Machine translation of natural languages, commonly known as MT, has multiple personalities. First of all, it is a venerable **scientific enterprise**, a component of the larger area of studies concerned with the studies of human language understanding capacity. Indeed, computer modeling of thought processes, memory and knowledge is an important component of certain areas of linguistics, philosophy, psychology, neuroscience and the field of artificial intelligence (AI) within computer science. MT promises the practitioners of these sciences empirical results that could be used for corroboration or refutation of a variety of hypotheses and theories.

But MT is also a **technological challenge** of the first order. It offers an opportunity for software designers and engineers to dabble in constructing very complex and large-scale non-numerical systems and for field computational linguists, an opportunity to test their understanding of the syntax and semantics of a variety of languages by encoding this vast, though rarely comprehensive, knowledge into a form suitable for processing by computer programs.

To complicate things even further, MT has a strong connection with the needs of modern societies. It has come to be understood as an **economic necessity**, considering that the growth of international communication keeps intensifying both at government (European Union, NAFTA, GATT) and business and commerce levels (exporters need product documentation in the languages of the countries where their products are marketed). The demand for faster and cheaper translations is strongóindeed governments, international agencies and industry keep supporting MT research and development, even though the return on their investment is, though not negligible, still rather modest.

Finally, MT is a gold mine for an amateur sociologist or a science reporter. This is one of the liveliest areas of computer science; its broad objectives can be readily understood by intelligent lay people. The practitioners of the field are rich sources of entertaining copy, as theyóknowingly or unknowinglyókeep staging a soap opera of inflated claims leading to dramatic disappointments, eminently quotable public debates and internescine fighting of almost religious proportions.

We will try to give the reader a coherent, if necessarily brief and incomplete, picture of the many ìfacesî of MT.

2. Is Machine Translation Impossible?

Many firmly established technologies and scientific theories, such as flying machines which are heavier than air or general relativity, had at some stage in their development been targets for purported disproofs. MT has been at this stage during all of its existence. Its established status as a technology is not new: the U.S. Federal Translation Division at Dayton, Ohio has been providing translations of Russian scientific publications to American scientists for decades at a rate of hundreds of thousands of words a day. The quality has not been high but sufficient for the scientistsí purposes, given that they know their discipline intimately. More recently, the European Commission has been using machine translation for rough but adequate translations of an increasing proportion of its internal memoranda between English and French, its two major languages. (Both these systems are based on SYSTRAN, the worldís oldest major MT system). So much for the reality of MT, and we shall produce empirical evidence in support of these claims later.

The proofs of the impossibility of MT ranged from philosophical considerations, often believed to derive from the anti-logical views of Wittgenstein who came to believe that there could be no abstract representation of the content of language, to more linguistic demonstrations. Among developed philosophical views, critics of MT often refer to the ìthe indeterminacy of translationî claims of Quine, the greatest living American philosopher. In the sceptical tradition of Hume, who did not question that there were causes and effects but asserted the connection between them could never be proved, Quine argued that one could never prove the equivalence of two sentences in different languages. He has been widely misunderstood on this point, since he did not intend to argue that there were no translations. Quine wrote ì...manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in no plausible sort of equivalence however looseî [1]. His main point is not that it is impossible to translateóhumans do it all the timeóbut that what is translated cannot reliably be shown to be the ìsame meaning.î Indeed, he was not attacking the notion of translation at all, but that of meaning in natural language. Since the task of machine translation is to reproduce what people do when they translate, at least as regards the output, Quineís position really poses no problem. As he wrote later: ìThe critique of meaning levelled by my thesis of the indeterminacy of translation is meant to clear away misconceptions, but the result is not nihilism. Translation is and remains indispensable. Indeterminacy means not that there is no acceptable translation but that there are many. A good manual of translation fits all checkpoints of verbal behavior, and what does not surface at any checkpoint can do no harmî [2].

Other theoretical positions that would render MT impossible include that of Whorf, the anthropologist, who argued that all meaning is culturally dependent, which implies that there cannot be translation between cultures with very different belief systems, norms and values.

Whorf explicitly extended this argument to scientific cultures, arguing that there could be no full translation of a sentence of physics into one of chemistry that described the ìsame situation,î a proposition that some readers might find perfectly acceptable. His own position was somewhat undermined by his printing, at the end of his best known book, a sentence-by-sentence translation into English of precisely the kind of American Indian language whose culture he deemed too remote for translation! The most telling and direct denial of MTóor more precisely of ìfully automatic high quality MTîócame from the Israeli philosopher Bar Hillel [3,) 4 (the latter paper was published in Advances in Computers)]. Bar Hillelís central claim was that fully automatic high-quality machine translation was unattainable with the technology of the times that could not guarantee word sense choice. His now-famous example was the following:

> Little John was looking for his toy box. Finally, he found it. **The box was in the pen**. John was very happy.

The word **pen** in the emphasized sentence has at least two meaningsóa writing pen and a play-pen. Bar Hillelís conclusion was that ìno existing or imaginable program will enable an electronic computer to determine that the word **pen** in the given sentence within the given context has the second of the above meanings.î

Later, Bar Hillel changed his position (see below) and conceded that MT had made great advances, but that this was not due to an improved theoretical basis, but a set of heuristics that seemed to aid programs:

> ìMT research should restrict itself, in my opinion, to the development of what I called before ìbags of tricksî and follow the general linguistic research only to such a degree as is necessary without losing itself in Utopian ideas.î [4]

It is interesting to note that Bar Hillelís position was not that far from Whorfís: both argued for the need for cultural knowledge in the translation process, and Bar Hillel, too, used scientific examples to make his point. In his claim that the computer must have some internal knowledge to do MT, Bar Hillel was very close to the defining position of the then nascent field of *artificial intelligence* and its program to model any distinctively human capacity that requires intelligence, which certainly would include translation. A sub-theme in AI is that such modeling requires ìcoded knowledgeî of the world. The difference between Bar Hillelís position and that of AI is that he thought it could not be done.

An important characteristic of Bar Hillelís position is that it straddles the often competing views of MT as science and as technology by suggesting that any means for resolving difficulties in language analysis and generation are appropriate, even if their theoretical nature is not yet under-stood by science. While paying homage to linguistics, most developers of MT systems in fact do not really use any of its findings in their system-building activities.

3. What Sort of Computation is MT?

We used the term ìheuristicsî above in connection with MT, and that term has often been taken as a tell-tale sign of a problem being a part of artificial intelligence, at least since John McCarthy declared AI to be the ìscience of heuristicsî [5]. The term is classically opposed in Computer Science to algorithms which, in the strong sense of the term, are computations which can be proved to terminate and produce provably correct results. Computations involving natural language are bad candidates for strong algorithms because there is no general agreement on the data to be ìprovedî by such algorithms: there is no general agreement on whether strings of words, in general, are or are not sentences of a given language. Even if Quineís arguments are not taken into account, there are no proofs that a given sentence in one language is a correct translation of a given sentence in another. There is a tradition in computational linguistics that still seeks algorithms, in a strong sense: namely the one descending from Chomskyís linguistic theories [6] and his original claim (since much modified) that the sentences of a language were a decidable set, in that an algorithm could be written to decide whether or not a given sentence string belonged to the set.

Linguistics-inspired theories of MT continue to be put forward, but never produce any substantial or realistic coverage of languages in empirical tests. The more interesting issue is the relationship of MT to its other natural parent discipline, AI. The main question is whether in order to perform successful MT, one needs to have a system which ìunderstandsî language to the same degree that is needed for intelligent robots. A tempting view is that it might be not necessary for the purposes of MT to understand the message in the text ìdeeply.î That many inferences necessary, for instance, for robotic response and action, are spurious for machine translation. Indeed, ambiguities of sense and reference can sometimes be ìbenignî in MT. Consider as an example the English sentence:

> The soldiers fired at the women and I saw several of them fall.

Do you see an ambiguity in it? Yes, ìthemî can be construed to refer either to ìwomenî or to ìsoldiers.î Of course, people have enough knowledge to make the connection between the event of shooting and its typical consequences. A simple machine translation program may not have such knowledge. But when we translate the sentence into Russian, we can just compose it of the translations of the component words directly:

—ÓÎ‰‡Ú˚ ‚˚ÒÚ□ÂÎËÎË ‚ ÊÂÌ˘ËÌ, Ë ˇ Û‚Ë‰ÂÎ, Í‡Í ÌÂÒÍÓÎ¸ÍÓ ËÁ ÌËı ÛÔ‡ÎË.
Soldiers fired at women and I saw how several of them fell

The problem is that when the sentence is translated into Spanish, the pronoun which translates ìthemî must be marked for gender. The correct gender can be only determined if we indeed know whether the soldiers or women fell. In the absence of evidence about any unusual events, we must conclude, based on our knowledge of the consequences of shooting, that it was some women who fell (this is an example of the so-called default reasoning, where we assume that had it been, in fact, the soldiers who fell, a special mention of this fact would have been present in the text because it is an unexpected situation and the authors could not rely on the readersí world knowledge). Worse, the machine translation system must be taught to deal with this kind of problem.

The debate about the depth of analysis will no doubt continue and different recommendations will be produced depending on the objectives of particular projects. It is important to realize that there are genuinely different theories of how to do MT, and only practical experiment will show which ones are right.

4. Main Paradigms for MTóDiverse Strategies for Solving or Neutralizing the Complexity of Language Use

Traditionally, MT systems were classified into three major typesó*direct, transfer* and *interlingua*. Direct MT systems rely on finding direct correspondences between source and target lexical units, and have been criticized for their *ad hoc* characteróthe apparent impossibility to generalize the translation rules; by 1970s, they have lost their scientific standing, though, notably, not their technological impact. Transfer systems involve a varying measure of target language-independent analysis of the source language. This analysis is usually syntactic, and its result allows substituting source language lexical units by target language lexical units, in context. Bilingual lexicons connect the lexical units of the source and target languages. Transfer systems permit taking into account syntactic sentence constituents in which lexical units appear. In interlingua systems the source language and the target language are never in direct contact. The processing in such systems has traditionally been understood to involve two major stages: (i) representing the meaning of a source language text in an artificial unambiguous formal language, the interlingua, and then (ii) expressing this meaning using the lexical units and syntactic constructions of the target language. Few large-scale interlingual systems have been fully implemented because of the very high complexity (both theoretical and empirical) of extracting a ìdeepî meaning from a natural language text.

In practice, essential differences between transfer-based and knowledge-based machine translation are still a subject of debate. The major distinction between the interlingua- and transfer-based systems is, in fact, not so much the presence or absence of a bilingual lexicon but rather the attitude toward comprehensive analysis of meaning. MT systems that do not strongly rely on ìdeepî text understanding, tend to prefer the transfer paradigm. Different transfer-based systems perform transfer at different levelsófrom simple ìbracketingî of sentence constituents to passing complex semantic properties of the input across the transfer link.

A recent trend in transfer-based machine translation is to downplay the need for structural transfer, that is, the stage of transforming standard syntactic structures of the source language into the corresponding target language structures. This trend is interlingual in nature. Transfer-based systems can also deal with lexical semantics; the language in which the meanings of source language lexical units are expressed is often the target language itself. This can be implemented through a bilingual lexicon featuring disambiguation information. The recent movement toward ìdeep transferî (see, e.g., work on Eurotra [7] is, in essence, a movement toward an interlingua architecture.

Distinctions between the transfer and the interlingua approaches are best drawn at a theoretical level. When practical systems are built, much of the work will be the same for both approaches, as a number of compromises are made in order to contain the amount of work necessary in preparing grammars, lexicons and processing rules. Some source language lexical units in an interlingua environment can, in fact, be treated in a transfer-like manner. Conversely, in those, very frequent, cases when there is no possibility of direct transfer of a lexical unit or a syntactic structure between two languages, a transfer system would benefit by trying to express the meaning of such units or structures in an interlingua-like fashion.

Most AI-oriented MT systems have been interlingual and have come to be known as knowledge-based MT (or KBMT) systems. Examples of interlingua systems that are not knowledge-based are CETA [8], DLT [9] and Rosetta [10]. The main difference between such systems and KBMT ones is the expected depth of source language analysis and the reliance of KBMT systems on explicit representation of world knowledge. The early roots of this movement were in Britain, with the work of Margaret Masterman and the Cambridge Language Research Unit, though similar contemporary developments can be traced in the USSR. The interlingua was assumed to be highly semantic/content-directed and not identical either with any interlingua of the sort that formal logic was traditionally taken as providing or with the syntax-driven interlingua of CETA.

Development of this style of work became significant within AI in the U.S. during the ë70s, particularly with the work of Schank and his school, and the work of Wilks. Schankís [11] MARGIE system took as input small English sentences, translated them into a semantic-network-based interlingua for verbs called Conceptual Dependency, massaged those structures with inference rules and gave output in German. In Wilksí system, called Preference Semantics, there was also an interlingua (based on 80 primitives in tree- and network-structures), between input in English and output in French, but there the emphasis was less on the nature of the representation than on the distinctive coherence algorithms (ìpreferences'î) for selecting the appropriate representation from among candidates. The reality and ubiquity of word-sense and structural ambiguity were a driving force behind that system. Both systems shared the assumption that traditional syntactic-based methods would not be able to solve that class of problems, and neither had separate syntactic components; the work of a syntactic component was performed under a semantic description. Again, both used MT only as a test-bed or application of more general claims about AI and natural language processing.

Another strand of MT work done in close association with AI has been that of Martin Kay at XEROX-PARC [11] Kay. He has emphasized the role of morphology, of machine-aided translation and of the structure of dictionaries on MT, but his most recent theme has been that of *functional grammar*, a formalism for syntax rules, usable in both analysis and generation, and hence part of the extensions of the linguistically-based movement in MT that began with GETA and TAUM, though now, once again a subject of independent interest in AI. The beginning of the revival of MT as a scientific discipline and an application of linguistic and computer technology must, however, be traced to the establishment of the Eurotra project and the MT efforts in Japan. Begun in 1978 and incorporating earlier European influences, such as the CETA project at the University of Grenoble, Eurotra was an ambitious, well-supported project aimed at providing MT capability among all official EEC languages At its peak, Eurotra employed about 160 researchers in a number of national groups and at the project headquarters in Luxembourg. The latest generation of Japanese MT efforts started around 1980, supported both by the government and industry, most notably with the Mu project at Kyoto University laying the foundation for the extensive Japanese industrial MT projects of the 1980s.

Further KBMT experiments were conducted by Jaime Carbonell, Richard Cullingford and

Anatole Gershman at Yale University [13] and Sergei Nirenburg, Victor Raskin and Allen Tucker at Colgate University [14]. Larger-scale development work followed, and a number of pilot KBMT systems have been implemented. The major efforts included ATLAS-II [15], PIVOT [16], ULTRA [17], KBMT-89 [18] and Pangloss. Some other systems (e.g., HICATS/JE [19] (Kaji 1988) are using some features of the knowledge-based approach (such as semantic primitives for organizing the dictionary structure) while still maintaining the overall transfer architecture.

Recently, the traditional classification of MT paradigms was modified by the resurgence of enthusiasm for the statistics-based approach to the problem.[1] Statistical MT explicitly rejects any kind of representation, whether AI or linguistics-oriented, whether morphological, syntactic or semantic. It relies on the availability of very large, aligned bilingual (or multilingual) corpora. The process of translation in this paradigm is, roughly, as follows: for each unit of SL input, find a match in the SL side of the bilingual corpus and write out in the output the unit in the TL side of the corpus which corresponds to the SL unit. The main issues in this paradigm are the quality and grain size of the alignment of the corpora and the problems of either lack of a complete match and of choosing the ìbestî match out of a set of candidates (for additional detail, see the discussion of the Candide project below).

Another statistics-inspired MT method has originated in Japan and come to be called example-based MT. It differs from the ìbasicî statistical method in that it relies on inexact matchingófor instance, matching a plural form of a phrase in input with a singular form of the same phrase in the corpus or matching a string where one (or more) words are substituted by their synonyms (*table* for *desk*) or even hyperonyms (*animal* for *dog*). Allowing inexact matching somewhat minimizes the need for huge bilingual corpora, as there is more chance for a ìhitî when an input phrase can match several corpus phrases. Example-based MT, thus, relies at least on morphological analysis and a thesaurus to establish the degree of matching between source text and the corpus. The metric for preferring some matches to others becomes very complex as more and more kinds of inexact matches are introduced.

4.1 Knowledge Sources for MT

Knowledge sources necessary for an MT system are determined, of course, by the approach used. For example, some statistics-oriented approaches often require the existence of large bilingual corpora. Some others, such as an experimental MT system at Kyoto University, need large lists of sample sentences against which a sentence to be translated is matched [21]. Rule-based MT systems however, make use of at least some, or possibly all of the following kinds of knowledge sources:

---

[1] The idea of statistical MT is not new: ìIt is well known that Western Languages are 50% redundant. Experiment shows that if an average person guesses the successive words in a completely unknown sentence he has to be told only half of them. Experiment shows that this also applies to guessing the successive word-ideas in a foreign language. How can this fact be used in machine translation?î [20]. In the 1950s, however, the state of the art in computer hardware and software precluded any serious experiments in statistical MT.

ï Morphology tables
ï Grammar rules
ï Lexicons
ï Representations of world knowledge

It is possible both to analyze and to represent the English language without the use of morphology tables, since it is inflected only to a small degree; for the analysis of a highly inflected language like Japanese, on the other hand, they are almost essential. Some analysis systems claim to use not an independent set of identifiable grammar rules, but they must somewhere contain information such as the fact that an article precedes a noun in English.[2] The third form of knowledge (lexical) appears in virtually every MT system, except for the purely statistical ones. And only KBMT systems claim to contain world knowledge representations. The distinction between the last two categories of knowledge can also be tricky: (in a German lexicon, for example, *das Fraulein* is marked as neuter in gender but, in the real world, it must be marked female, as the word means ìa young womanî). We should deduce from this that a lexicon is a typically a rag-bag of information, containing more than just *semantic* information about meanings.

5. The Evolution of MT Over its 50-year History

Much of the extant scepticism about MT is outdated and comes from sociological factors, notably including the early enthusiasm for the technology, followed by its virtual collapse in the U.S. in mid-1960. Was early MT so bad as everyone now says? Was the information theory and its underlying cryptology-inspired decoding hypothesis absurd? A brief historical discussion can help answer these questions.

It is customary to consider the so-called Weaver memorandum as the starting point of research in MT.[3] In 1949 Warren Weaver, then a vice president of the Rockefeller Foundation, distributed 200 copies of a letter in which he suggested the concept of MT to some of the people who were likely to have an interest in its development. Even though the memorandum was predominantly a strategic document, several important theoretical and methodological issues were discussed, including the problem of multiple meanings of words and phrases, the logical basis of language, the influence of cryptography and the need to analyze language universals. The memorandum aroused significant scientific and public interest. In 1948, a University of London team led by Andrew Booth and Richard Richens was the worldís only MT research and experimentation center. In the two years after the Weaver memorandum, MT work started in the U.S. at the

2 Although there is room for doubt as to which of the first two categories (morphology or grammar) certain items of linguistic knowledge belong to (in Italian, for example, forms such as pronouns may stand alone but can also act as suffixes to verbs: e.g. *daglielo*), in Japanese and English this ambiguity of type is very unlikely indeed.

3 Claims about the actual starting point of a field of study are notoriously inexact. Some work on mechanical translation preceded the Weaver memorandum. Not only did Booth and his colleagues start such work in Britain in about 1946, but several patent applications were filed in the 1930s for translation devices, including very ingenious ones by the Soviet engineer Smirnov-Troyansky [22].

Massachusetts Institute of Technology, the University of Washington, the University of California at Los Angeles, the RAND Corporation, the National Bureau of Standards, Harvard University and Georgetown University.

The major concepts, topics and processes of MTósuch as morphological and syntactic analysis, pre- and post-editing, homograph resolution, interlingual representation of text meaning, work in restricted vocabularies, automating dictionary look-up, and so onówere first defined and debated at that time. The first scientific conference on MT was held in 1952 at MIT, and the first public demonstration of a translation program took place in 1954. This was the famous Georgetown experiment which involved translating about 50 Russian sentences, preselected from texts on chemistry, into English.) This experiment was perceived by the general public and sponsors of scientific research as strong evidence for the feasibility of MT. The wide publicity and resonance of this experiment has also led to the establishment of MT projects outside the U.S., notably in the Soviet Union.

Through the 1950s and into the following decade, MT research in the U.S. and the Soviet Union continued and grew. Attempting to scale upwards from the initial Georgetown experiment, however, proved more difficult than expected, as translation quality declined with expanded coverage. The quality of fully automatic translation remained largely below an acceptable level and required extensive human post-editing, as can be seen from the following example, an excerpt from the output of a 1962 demonstration of the Georgetown GAT system, one of the best examples of MT from Russian to English:

> By by one from the first practical applications of logical capabilities of machines was their utilization for the translation of texts from an one tongue on other. Linguistic differences represent the serious hindrance on a way for the development of cultural, social, political, and scientific connections between nations. Automation of the process of a translation, the application of machines, with a help which possible to effect a translation without a knowledge of a corresponding foreign tongue, would be by an important step forward in the decision of this problem. [23] (Hutchins quoting Dostert 1963)

Still, researchers in MT remained largely optimistic about the prospects of the field. ìThe translation machine...,î wrote Emile Delavenay in 1960, ìis now on our doorstep. In order to set it to work, it remains to complete the exploration of linguistic data.î Two factors brought an end to the first enthusiastic period of MT research and developmentóthe so-called ALPAC report and the persistently intractable problems of treatment of meaning (word sense choice and semantic structure issues that we have already touched on).

Since Bar Hillel was one of the early champions of MT and had intimate knowledge of the research in the field, his critique (and ìdisproofî described above) has had a wide resonance in public attitudes toward MT, as well as among its sponsors in the U.S. government and industry. Coupled with the increased difficulty of problems facing MT research after the initial successes, and notwithstanding the fact that many of the then-current projects (notably, at Georgetown University and IBM) pursued exactly the type of MT research recommended by Bar Hillelónamely, a combination of machine translation with human post-editingóthis criticism started the process of reassessment of attitudes toward the field.

The reassessment culminated in the publication in 1966 of the National Academy of Sciences Automatic Language Processing Advisory Committee (ALPAC) report which was critical of the state of research in MT and recommended reductions in the level of support for it. Remarkably, the central ALPAC argument was purely economicóthe cost of machine translation, with human post-editing, was at the time higher than the cost of purely human translation. Thus, it was a judgment of MT as technology. Notably, no negative judgment was pronounced on MT as a scientific enterprise.

The principal mistake of the early MT workers was that of judgment: the complexity of the problem of natural language understanding was underestimated. The variety and the sheer amount of knowledge necessary for any solution to this problem proved to be enormous, so that the success of MT as technology became dependent on the solution to the problem of knowledge acquisition and integration. It would take more than 15 years for MT to start a scientific comeback in the U.S.

While the ALPAC report reduced American efforts in MT,[4] research and development continued to grow in several scientific groups in the Soviet Union, Canada, Germany, France and Italy, as well as in a small number of commercial institutions in the U.S. Notable MT achievements in the 15 years after the ALPAC report included the development and everyday use of the first unquestionably successful MT system, TAUM-METEO, developed at the University of MontrÈal and used routinely since 1977 to translate weather reports from English into French. The MT program SYSTRAN was used during the Apollo-Soyuz space mission in 1975, and in the following year was officially adopted as a translation tool of the European Economic Community.
MT research in the U.S. gradually revived. MT activities at various scientific meetings have significantly intensified, and several conference series devoted specifically to MT have been founded in the past decade. 1999 has seen the Seventh International MT Summit conference and the Eighth International Conference on Theoretical and Methodological issues in MT. New research groups have been set up all over the world. The general mood of the conferences reflects a new optimism based on modern scientific advances and the fact that the need for MT in the 1990s is vastly more pressing than in the world of 40 years ago.

5.1 MT ìGenerationsî

The early division of MT systems into direct, transfer and interlingua was somewhat parallel to the recognition of several chronological ìgenerationsî of MT work.[5] The first generation was prevalent before the ALPAC report, when all systems were direct. In the wake of the report, attempts were made to program systems with rules closer to systems of contemporary syntactic theory. These included GETA [24], a system at Grenoble, France, due to the late Bernard Vau-

---

4 though far from made them extinct, as is a popular myth

5 Though generalizations such as this are notable for an abundance of special cases and exceptions, a brief discussion should give the reader a better understanding of the field of MT. One must be very careful with using the term ìgenerationî: as with ìN-th generation computingî its role is essentially rhetorical rather than descriptive and used to claim novelty for the product of the speaker.

quois, and based on a version of valency theory; the TAUM system was a descendent of GETA. Such systems relied on tree-structure representations that allowed complex structures to be attached to nodes and, in that sense, were, in fact, richer than those then available in syntactic theory in the Chomskyan tradition. This type of direction, begun in the late 1960s, is often referred to as ìsecond generationî MT.[6] The ìthird generationî MT has come to be associated with knowledge-based approaches. It started in earnest at the same time as the ìsecond generationî and blossomed in the 1980s. While a brief flirtation with connectionist approaches (which were based on processing language using artificial neural networks, electronic models of the human brain) does not constitute a significant enough trend, the flowering of statistical MT, roughly after 1990, should be recognized as MTís ìfourth generation.î[7]

6. Choices And Arguments For and Against MT Paradigms?

6.1 Putting a Natural Language in the Center

Some MT researchers maintain that there is no need to use an artificial language for representing meanings, that a natural language, such as Aymara or Sanskrit, will do. Others maintain that instead of inventing new artificial languages, MT can use some of the available artificial languages, such as Esperanto. Such claims are driven by two considerations: (i) that it is difficult to design and actually acquire the syntax and the semantics for an artificial language; and (ii) that some natural languages exhibit a sufficiently ìlogicalî character to be used directly by computers. While the latter consideration is, in the final analysis, romantic or propagandist (ìthe language I like the most is the most logicalî) and not scientific, the former claim is only too true. It is, therefore, natural, to continue to look for alternatives to complex artificial interlinguas. Such alternatives, however, are not easy to find. Indeed, the crucial difference between languages used by humans and languages designed for the use of a computer program is the mechanism which is expected to process them. Natural languages are used by humans. Artificial languages have computer programs as understanding agents. For excellent understanding machines such as humans *brevity* is at a premium, even at the expense of ambiguity and implicitness of some information.

---

[6] Such usage is full of historical irony, in that, for example, the broad theoretical basis of ìsecond generation MTîóthe adequacy of a family of phrase structure grammars for MT and natural-language processing generallyóstrongly predates some current developments in AI and computational linguistics, in which a resurrected form of phrase structure grammar has seized the center stage from more semantic and knowledge-based methods in natural language processing. Later work following in this tradition of using more perspicuous and context-free syntax rules for analysis in MT included Melbyís work on ìjunction grammarî at Brigham Young University [25], and Slocumís METAL system at Austin, Texas [26]. A later addition was the EUROTRA system, developed for the European Community in Luxembourg between 1982 and 1992 [27]. This attempted initially to blend a GETA-like syntax with some of the insights from AI-based natural-language understanding, at its semantic levels. However, this was eventually abandoned, and the project turned to a variant of definite clause grammar, which is to say, in MT-historical, rather than AI-historical, terms, firmly entrenched itself within second-generation techniques.

[7] A principal and misleading feature of the ìgenerationî analogy, when applied to MT, is that it suggests successive time segments into which the different methods fall. As already seen, that can be highly misleading: the SYSTRAN system, for example, is a surviving form of a ìfirst generationî system, existing alongside second, third and later, generational developments. Evolutionary phylae would be a much better metaphor here, because earlier systems, like sharks, survive perfectly well alongside later developments, like fish and whales.

For computer programs, lack of ambiguity and explicitness of representation is at a premium, at the expense of verbosity. Thus, the key to an effective interlingua format is that it be unambiguous and explicit.

Furthermore, the characteristics of the communication channel suggest that the texts in languages spoken by humans be single-dimensional strings. With computers, knowledge can have a much more complex topologyóof hierarchies or even multidimensional lattices. The only way in which one can say, loosely, that a natural language is used as interlingua is when lexical units of this language are used to tag ontological concepts. However, there is ample additional representational apparatus which is entailed in designing an interlingua. It is not impossible to use lexical units from certain natural languages or human-oriented artificial languages like Esperanto as markers for ontological concepts. In fact, in our own work we use combinations of English lexemes to tag concepts. However, in order to turn such a language into an efficient text meaning representation for MT (at least some of) these meanings will have to be explicated in terms of their properties and typical connections with other concepts.

6.2 Can One Avoid Treatment of Meaning?

The search of ways of avoiding the need for the use of knowledge and ìdeepî analysis of text takes other forms, too. That meaning understanding is not necessary is also maintained by those who observe that the polysemous Spanish noun *centro* is translated into German as *zentrum* no matter which of the senses of *centro* was used in the original text. The question is then posed: ìWhy waste time detecting and representing the meaning of the input string when the target language correlate is always the same?î Similar claims have been made about syntactic ambiguities [28] and ambiguities of prepositional phrase attachment [29]. A typical formulation of this position is given by Ben Ari *et al.* [30]:

> ìIt must be kept in mind that the translation process does not necessarily require full understanding of the text. Many ambiguities may be preserved during translation [Pericliev 84], and thus should not be presented to the user (human translator) for resolution.î

Similarly, Isabelle and Bourbeau [31] contend that:

> ìSometimes, it is possible to ignore certain ambiguities, in the hope that the same ambiguities will carry over in translation. This is particularly true in systems like TAUM-aviation that deal with only one pair of closely related languages within a severely restricted subdomain. The difficult problem of prepositional phrase attachment, for example, is frequently bypassed in this way. Generally speaking, however, analysis is aimed at producing an unambiguous intermediate representation.î

6.3 More ìDisproofsî for KBMT?

Some MT researchers adopt the position that different languages employ different concepts, or employ concepts differently, and this short-circuits attempts at meaning extraction. Thus Amano [32] writes: ìNatural languages have their own articulation of concepts according to their culture. Interlingua must naturally take account of this.î To illustrate this point, Amano reports that where the English word moustache is customarily defined in English dictionaries as comprising hair on the upper lip, the Japanese *kuchi-hige* is defined in one (unspecified) Japanese dictionary as a ìbeard under the nose.î (Actually, the Kanji ideographs for kuchi-hige stand for ìlipî or ìmouthî plus ìwhiskers.î) From this we are urged to infer that what Japanese speakers mean by *kuchi-hige* is somehow different than what English speakers mean by *moustache*. Of course, this opinion is simply a particularly hirsute version of Sapir-Whorfism that depends crucially on the vagaries of dictionary entries. In addition, the claim displays a common misunderstanding of the concept of interlingua. What differs among languages is not the meaning representation but rather the lexical and syntactic means of realizing this meaning. The meaning of *kuchi-hige* and *moustache* will be represented in the same way in an interlingua text. The realizations of this meaning in the two languages will be different. It is in the interlingua-TL dictionary that a connection is established between an interlingual meaning representation and the language-particular linguistic expression.

This is not the place to argue against linguistic and cognitive relativism. The idea of linguistic relativity is, in fact, neutral with respect to the tasks of computational linguistics. It should be sufficient to point out that however efficient dictionaries might be as explicators of meaning for humans, it is a mistake to appeal to them as formal indices of a cultureís conceptual structure. To contend that meaning exists within a language but not across languages means to subscribe to an extreme sort of relativism usually associated with treating language as a mass of individual dialects or even idiolects. In practice, of course, indigenous realia can be described encyclopedically and then assigned a linguistic sign (possibly, a direct calque from the original language). A slightly different ìcultural-imperialistî argument against language-independent meaning representation states that the way an interlingua is built reflects the world view behind one dominant language. Examples of phenomena with respect to which ìcultural imperialismî can be established include the cross-linguistic difference in subcategorization behavior of verbs, the grain size of concept description and the difference in attitude similar to the moustache example above. For instance, a single interlingua concept can be suggested to represent the main sense of the English put (as in *put a book/glass on the table.* This might be considered a case of English cultural imperialism because in Russian this meaning can be expressed either as *polozití* or *postavití* depending on some properties of the object of put.The difference can be glossed as that between put flat and put upright. A book can be ìputî either way; a glass will be usually ìput upright.î

6.4 Ideal Interlingua and Its Practical Realizations

The view of the interlingua as a representation capturing all meanings in all languages is too simplistic because it talks about an ideal case. As Makoto Nagao [33] put it:

> ì...when the pivot language method [i.e, interlingua] is used, the results of the analytic stage must be in a form which can be utilized by all of the different languages into which translation is to take place This level of subtlety is a practical impossibility.î

On a more technological level, [34] justifies the choice of the transfer approach in the Metal project as follows:

> ìMETAL employs a modified transfer approach rather than an interlingua. If a meta-language [an interlingua] were to be used for translation purposes it would need to incorporate all possible features of many languages. That would not only be an endless task but probably a fruitless one as well. Such a system would soon become unmanageable and perhaps collapse under its own weight.î

This ìmaximalistî view of interlingua is so popular probably because it is conceptually the simplest. In operational terms, however, it is not more complex to conceive such an interlingua as a set of bilingual dictionaries among all the language pairs in the world. A practical interlingua should be viewed both as an object and as a process. Viewed as an object developed in a concrete project, an interlingua should be judged by the quality of the translations that it supports between all the languages for which the corresponding SL-interlingua and interlingua-TL dictionaries have been built. As a process, its success should be judged in terms of the ease with which new concepts can be added to it and existing concepts modified in view of new evidence.

6.5 Is an Interlingua a Natural Language in Disguise?

The persistent problem for those in MT who work with interlinguas in particular and knowledge representations in general is to explain why symbolic internal structures always look so much like a real language when set down. It is widely supposed that knowledge-based machine translation requires grounding in a fully interpreted logical calculus, that a meaning-based approach cannot be presented with such formal rigor and hence that meaning-based MT cannot succeed. This argument may be understood as demanding formal proofs of the correctness of meaning representations. Without such proofs, it is supposed, there is no guarantee that a translation will be free of contradiction or that the same meanings will be always represented similarly.

This formalist approach to machine translation stems from a logic and philosophy of language tradition which tends believe that there is no distinction in principle between natural and formal languages. But even if this supposition were correct, it would not follow that uniformly formal representations are necessary for the task of machine translation. As Wilks [35] put it:

> [] we do need representations [], but their form, if interpretable, is largely arbitrary, and we may be confident it has little relation to logic. I shall restate the view that the key contribution of AI in unraveling how such complex tasks as ìunderstandingî might be simulated by a machine lies not in representations at all but in particular kinds of procedures [] It would be the most extraordinary coincidence, cultural, evolutionary, and intellectual, if what was needed for the computational task should turn out to be formal logic, a structure derived for something else entirely.

The demand for proof that a target language text will contain no contradiction is of course a demand that cannot be met. But, fortunately, the problem of avoiding contradictionóin machine translation in particular and natural language processing in generalóis an empirical issue and not clearly delimited by formalist claims and purported requirements. That is to say, while it might be nice to be able to offer such proofs, it would be a grievous error to abandon any enterprise unable to provide a formal proof of its future success. Indeed, the formalist gambit has been tried against any number of sciences, including physics, and has come up short. Human translations are not ìprovably correct.î Moreover, very few computer programs can be actually proved correctóand then only with respect to formal specifications and not to real-world implementations.

6.6 The Ideal and the Reality of Statistical MT

The attractiveness of the statistical approach to MT centers on the claim that it can perform MT without a glimmering of understanding of the organization of language or even the actual languages involved! In essence, the ideal, purely statistical method was described by the IBM researchers [36] is an adaptation of one that worked well for speech decoding [37]. The method, which was claimed to underlie the French-English translation system Candide, establishes three components: (a) a trigram model of English sequences; (b) the same for French; (c) a model of quantitative correspondence of the parts of aligned sentences between French and English. In the Candide project, the first two are established from very large monolingual corpora in the two languages, of the order of 100 million words, the third from a corpus of *aligned* sentences in a parallel French-English corpus that are translations of each other. All three were provided by a large machine-readable subset of the French-English parallel corpus of Canadian parliamentary proceedings. Both (a) and (b) are valuable independent of the language pair and could be used in other pairings (which is why Candide claims now to be a *transfer* MT project).

In very rough simplification, the Candide system works as follows: an English sentence yields likeliest equivalences for word strings (substrings of the English input sentence) i.e. French word strings. The trigram model for French rearranges these into the most likely order, which is the output French sentence. One of their most striking demonstrations is that the trigram model for French (or English) reliably produces (as the likeliest order for the components) the correct ordering of items for a sentence of 10 words or less.

What should be emphasized is the enormous amount of pre-computation that this method requires to produce alignments and the trigram analysis. But even after all the pre-computation is completed, a ten-word input sentence required an additional hour of computation to produce a translation. This figure will be undoubtedly reduced with time and hardware expansion but it gives some idea of the computational intensity of the Candide method.

Because of this computational complexity, Candide has, in fact, taken in whatever linguistics has helped: morphology tables, sense tagging (which is directional and dependent of the properties of French in particular), a transfer architecture with an intermediate representation and an actual or proposed use of bilingual dictionaries. The pure, ideal statistical MT has been shown to failóthe purely statistical results topped out at around 40% of sentences acceptably translated. Fortified with rule-based addenda, Candide reached much higher acceptability figures. The Candide story is, indeed, one of the triumphs of the technology-oriented, ìbag of tricksî mind set over the ìpureî scientific attitudes.

6.7 Statistical MT Does Not Solve All Hard Problems

There remain crucial classes of cases that seem to need symbolic inference. Consider the following example:

> *PRIEST* IS CHARGED WITH POPE ATTACK (Lisbon, May 14) *A Spanish priest* was charged here today with attempting to murder the Pope. *Juan Fernandez Krohn*, aged 32, was arrested after *a man armed with a bayonet* approached the Pope while he was saying prayers at Fatima on Wednesday night. According to the police, *Fernandez* told the investigators today *he* trained for the past six months for the assault. *He* was alleged to have claimed the Pope ëlooked furiousí on hearing *the priestís* criticism of his handling of the churchís affairs. If found guilty, *the Spaniard* faces a prison sentence of 15-20 years. (The Times,15 May 1982)

The emphasized phrases all refer to the same man, a vital fact for a translator to know since some of those phrases could not be used in any literal manner in another language (e.g. ìthe Spaniardî could not be translated word-for-word into Spanish or Russian). It is hard to imagine multiple identity of reference like that having *any* determinable statistical basis.

Other, more general, reasons to question the promise of pure statistical methods are self evident. Statistical information retrieval (IR) generally works below the 80% threshold, and the precision/recall trade-off seems a barrier to greater success by those methods. Yet it is, by general agreement, an easier task than MT and has been systematically worked on for over 35 years, unlike statistical MT whose career has been intermittent. The relationship of MT to IR is rather like that of sentence parsers to sentence recognizers. A key point to note is how rapid the early successes of information retrieval were, and how slow the optimization of those techniques has been since then! The standard model of a single language in statistical processing is a trigram model because moving up to even one item longer (i.e. a tetragram model) would be computationally prohibitive. This alone must imposes a strong constraint on how well the pure statistical approach can do in the end, since it is universally agreed that any language has phenomena that ìconnectî outside the three item window. The issue is how far one can get with the simple trigram model (and, as we have seen, it yields a basic 40%), and how well distance phenomena in syntax can be finessed by various forms of information caching.

7. MT in the Real World

The purpose of machine translation is, on the one hand, to relieve the human translator of the need to work with tedious, repetitive and aesthetically unsatisfying material and, on the other, to speed up and facilitate worldwide information dissemination. Translating fiction and poetry is not a task for machine translation. The focus is on translating technical texts and other expository texts.

It is hardly necessary to argue for the fundamental importance of the ability to disseminate information across linguistic borders. The volume of such information is already very large and is bound to increase with the growth of international trade and international political cooperation. A study by the Japan Electronic Industry Development Association estimates the size of the translation market in Japan in 1988 as being close to a trillion yen (about $8 billion) annually, calculated as about 240-million pages of translation at 4,000 yen per page, a page containing 400 characters of kana or 125 English words. The market was expected to increase about twofold in the following two years.

The pressure is growing for MT researchers and developers to come up with products that will ìbreak through.î What, then, should MT researchers do? As fully automatic, high-quality MT in large domains is not feasible, ìsuccessfulî MT should be redefined. This could be done by any combination of the following methods: (i) accepting lower-quality output, (ii) making MT only partially automated, and (iii) making sure that the texts selected for translation can, in fact, be automatically processed by the system.

7.1 Varying the Acceptability Threshold

In many cases, especially when a machine-translated text is not intended for publication, the quality standards for MT can be effectively lowered without tangible harm. This situation arises, for instance, when an expert in a scientific field would like to read an article in his or her field of expertise but that is in a language he or she does not know. In the best case, a medium-quality machine-translated text will suffice for understanding the content. At worst, a decision could be made on the basis of such a translation whether the article is interesting enough for it to be translated by other means, such as a more costly human translation. Intermediate positions, such as human post-editing of the rough translation may prove economically superior to a complete retranslation.

7.2 Partial Automation in MT

Until very recently human intervention in the MT cycle came predominantly in the form of post-editingóa human editor improving the results of MT before it is submitted to publication or otherwise disseminated. Among such systems are SYSTRAN and its direct predecessor, GAT, SPANAM, HICATS and some others. A major objective in systems relying on post-editing is to make human involvement as painless as possible in comparing source and target texts and correcting any translation errors. Many professional post-editors report serious difficulties in improving machine-translated output compared to that of human translators. If MT is viewed in purely economic terms, as a technology, then developing interactive computer environments (ìTranslatorís Workstationsî) which facilitate the work of a post-editor is as important a task as building ìmainlineî MT systems.

7.3 Restricting the Ambiguity of Source Text

The most straightforward way of restricting the source text ambiguity is by choosing a sufficiently narrow subject domain. The terminological inventory in such a domain will be limited and relatively easy to describe. This approach to MT has come to be known as sublanguage-oriented MT. The best example of the sublanguage approach is the operational MT system TAUM-METEO, developed at the University of MontrÈal and delivered to the Canadian Weather Service for everyday routine translations of weather reports from English into French. The system operates very successfully, practically without human intervention. Its vocabulary consists of about 1,500 items, about half of which are place names. There is very little lexical ambiguity in the system because words are expected to be used in only one of their sensesónamely, the one that belongs to the subworld of weather phenomena. For instance, the word front in TAUM-METEO will be understood unequivocally as a weather front.

Finding well-delineated, self-sufficient and useful sublanguages is a very difficult task, and this is one of the reasons why the success of TAUM-METEO has not yet been repeated by any other operational system. Unfortunately, most subworlds (e.g., computer manuals, finance or chemistry) are not as restricted with respect to vocabulary size and syntactic diversity as repetitive weather forecasts.

As an alternative, texts can be deliberately simplified for the use of an MT system. In order to prepare texts for machine translation a human pre-editor may be employed who reads the input text and modifies it to include only the words and constructions which MT system is able to process it automatically. Difficult and overly ambiguous words and phrases are replaced with those the editor knows that the program will be able to handle. A version of this method has been widely employed in industry (e.g., the TITUS system, the Xerox Corp. Docutran, etc.). Like post-editors, pre-editors must also be supplied with interfaces and tools.

7.4 Statistical MT and the Economics of Corpora

In one sense, what the Candide project has done is partially automate the construction process of traditional MT systems, such as SYSTRAN: replacing laborious error feedback with statistical surveys and lexicon construction. However, Candide is, unlike other projects, totally tied to the bilingual corpus, the Rosetta Stone, one might say, of statistics-oriented MT. We should remember, too, that their notion of word sense is only and exactly that of correspondences between different languages, a wholly unintuitive one for many people. The problem for statistical MT is that few vast bilingual corpora are available in languages for which MT is needed. If, however, they had to be constructed by hand, then the economics of what Candide has done would change radically. By bad luck, the languages for which such corpora are available are also languages in which traditional MT systems (e.g., SYSTRAN) already have done pretty well, so for statistical MT to be taken seriously as a practical force, it will have to overtake, then widen the gap with, SYSTRANís performance. They may be clever enough to make do with less than the current 100-million-word corpora per language, but one would naturally expect quality to decline as they did so. This resource argument could be very important: it has been shown[8] for the case of statistical taggers, that any move to improve the performance by adding more higher-level knowledge structures always ended up requiring a much larger corpus than expected. This observation contravenes the claims by MT statisticians, so an important point to watch in the future is to monitor the availability of adequate bilingual corpora for the domain-specialized MT that is most in demand (such as airline reservations or bank billings): Hansard is large but is very general indeed.

7.5 Novel Applications I: Translation of Spoken Language

MTís sister area, speech processing, has progressed significantly over the past 20 years. Improvement in speech recognition and synthesis systems has whetted the appetite for practical

8 by the British computational linguist Geoffrey Leech

applications involving interpretationótranslation of spoken text. There are a number of uses for such technologyófrom multilingual telephony to personal communication. A number of projects have been devoted to this problem, which compounds the complexity of MT with the considerable complexity of the speech recognition problem as such. A variety of relatively small-scale efforts have been sponsored by the U.S. government, while much more ambitious projects have been put together in Japan (the ATR Laboratory) and Germany (the Verbmobil project). The goals of such projects are necessarily modest. However, several spectacular demonstrations of speech-to-speech translation have been staged over the past three years, for instance, that of the Spanish-to-English system at AT&T Bell Laboratories or the tripartite English-German-Japanese demonstration by ATR, the Siemens Nixdorf Corporation of Germany and Carnegie Mellon University in Pittsburgh. Connected by a communication satellite, three operators in Munich, Kyoto and Pittsburgh were able to communicate speaking in their own language, while the translation systems generated translations of their interlocutorsí utterances into the native tongues of the recipient. The conversation had to do with inquiries about registration for a scientific conference.

This latter demonstration has been, indeed, a spectacular technological achievement, especially with respect to integrating a variety of component systemsólanguage processors, speech recognizers, speech synthesizers, communication interfaces, etc. However, the envelope of the speech translation technology was arguably not pushed very hard by this demonstration. That this achievement was reported in national newscasts in Japan, Germany and the U.S. and on the front page of *The New York Times* may have been a mixed blessing, as this coverage may have unduly raised the expectations of a variety of parties strongly interested in developing this technology. In fact, the demonstrated system operated on a very small vocabulary of less than 500 lexical units in each language, and the prospects for rapid scaling-up of the size of vocabulary used by this demonstration are at best not immediate. The feeling of *deja vu* was experienced strongly by everybody who remembers the situation caused by the Georgetown experiment. Speech-to-speech MT must not ìoversellî its achievements and promise and thus avoid repeating a crucial mistake of the early MT. It is not wise to risk an ALPAC report on speech translation.

7.6 Novel Applications II: Multi-Engine MT

Current MT projectsóboth ``pure'' and hybrid, both predominantly technology-oriented and scientific are single-engine projects, capable of one particular type of source text analysis, one particular method of finding target language correspondences for source language elements and one prescribed method of generating the target text. While such projects can be quite useful, we believe that it is time to make the next step in the design of MT systems and to move toward adaptive, multiple-engine systems.

Practical MT systems are typically developed for a particular text type (e.g., weather reports, financial news articles, scientific abstracts) and for a particular end useóe.g., assimilation or dissemination of information. Special cases, such as translating an updated version of a previously translated text, abound in the real-world practice. Gains in output quality and efficiency can be expected if a machine translation environment can be made to adapt to a task profile. Thus, for example, for translating abstracts of scientific articles in order to select just those that are of particular interest to a customer, a statistics-based approach might be most appropriate.

Example-based translation seems to be most promising for translating new versions of previously translated documents. This correspondence between technique, input text type and end use (or output text type) provides further motivation for moving toward adaptive, multiple-engine systems.

Two approaches to adaptivity in MT have been formulated. Both presuppose an MT environment in which a number of MT engines are presentófor instance, one (or more!) each of knowledge-based, statistics-based, transfer-based or example-based engines can be used. In one of the approaches all available engines are ìunleashedî on an input text and the final output is assembled from the best text segments, irrespective of which engine produced them. In another approach a heuristic ìdispatcherî decides which of the available engines holds the highest promise for a given input text and then assigns the job to that engine.

The former approach involves more processing but allows an *a posteriori* selection of the best results. The latter approach saves cycles but relies on heuristic *a priori* selection of the best output. In this latter case, the quality of the heuristics for the dispatcher module is crucial, but additionally, the approach expects each of the component engines to be of rather high quality, since they would not (as is the case in the other approach) be ìbailed outî by other engines in case of failure.

7.7 Novel Applications III: MT and Other Kinds of Text Processing

MT has recently lost its hitherto uncontested hold on multilingual text processing. For several years there has been an effort, supported by the U.S. government, to develop systems for information retrieval from text in multiple languages. Several such programs (MUC, Tipster, TREC, etc.) have attracted a number of participants in the academia and industry. These researchers have based their systems on a variety of basic approaches (including, very prominently, the venerable ìbags of tricksî approach). However, and quite naturally, none of the systems built to extract a pre-defined set of information elements from texts in English or Japanese, to say nothing about systems devoted to filtering message traffic by directing the messages into a variety of topically organized ìbins,î have opted to use a full-scale machine translation system in their processing. This decision is natural because it has been widely perceived that the tasks of information extraction and message routing can be performed without the need for a ìblanketî coverage of the input, as is necessary in MT. MT was perceived as too heavy and too unreliable a weapon for this task. So, in order to build an information retrieval system in two languages, the developers decided to develop two bags of tricks, one for each language involved.

The information retrieval task has met with some success in the current formulation. At the same time, it is possible to argue that this task is an attractive application for machine translation. Indeed, why not have just one bag of information retrieval tricks, oriented at, say, English, and use it both on the texts that were supplied in the original English and texts that a battery of MT systems have produced from inputs in a variety of languages. Granted, the output of the translation programs should be expected to be less than high-quality. But the same argument can be invoked here that was used in deciding not to use MT for information retrieval: the latter can be done without a ìblanketî coverage of the text, without succeeding in analyzing (or even attempting to analyze) each and every input word. Good information retrieval may result from less than high-quality translation results.

8. The Current Situation

MT technology today is still immature in many senses:

ï it is too dependent on the task, i.e., the carefully pre-specified kinds of texts;

ï the effort to acquire sufficient knowledge for broad-domain fully automatic translation is still too high;

ï text analysis techniques still concentrate predominantly on the syntax and semantics of single sentences, without sufficient coverage of the phenomena of multi-sentence meaning;

ï technology is brittle in the face of unexpected input due to:
-- words missing from the lexicons
-- word and phrase meanings outside of the set of senses in the lexicon
-- input errors;

ï systems are still too dependent on a particular pair of source and target languages; and

ï no effective and general reference treatment mechanisms have been proposed.

At the same time, the intensity and variety of MT research is arguably at its all-time high. Work is being carried out in most sectors of the following opposition continua:

ï fully-automatic/machine-assisted (a whole range of sophistication of machine assistance, from simple post-editing to the use of an advanced TWS);

ï high-quality output/rough translations;

ï language-pair-tuned/multilingual;

ï unconstrained input/onstrained input (possibly, pre-editing; authoring systems to assist in production of legal constrained text);

ï text/speech/combination thereof;

ï complete translation/abstracting (summarization)/DB record creation;

ï multilingual generation from text/tabular info/DB records/graphics;

ï rule-based/corpus-based MT;

ï interlingua-oriented/transfer-oriented/specialized ìblanket coverage of small domainî glossary-based translation; and

ï full-blown statistical MT/example-based MT.

9. Conclusion

The new optimism of MT researchers and sponsors is based on spectacular advances in computer technology (drastic improvements in processing speed and memory capacity of computers, advances in computer architecture, emergence of database technology, development of high-level programming languages and interactive programming environments, etc.) and computational linguistics (in particular, techniques for morphological and syntactic analysis and synthesis of natural language texts). Advances in automatic processing of meaning and techniques of human-computer interaction are also an important component of the current MT paradigms. With the knowledge of the past difficulties, and, therefore, with a realistic assessment of the possibilities of MT technology application, the current MT projects are well equipped to produce a new wave of scientifically sound and practically useful machine translation systems. These systems are being designed both to compete and to cooperate with humans in translating a wide variety of scientific, industrial, official, journalistic and other texts. And, most significantly, low-cost, rapid-response MT systems promise to increase the volume of timely translations manyfold, addressing a societal need for timely delivery of currently untranslated quality information whose source may be in a number of different languages.

10. References

1. Quine, W.V. 1960. Word and Object, MIT Press.

2. Quine, W.V. 1987. Quiddities: an intermittently philosophical dictionary/W.V.Quine, Cambridge, Mass.: Belknap Press of Harvard University Press.

3. Bar Hillel, Y. 1959. Report on the state of machine translation in the United States and Great Britain. Technical report, 15 February 1959. Jerusalem: Hebrew University.

4. Bar Hillel, Y. 1960. The Present Status of Automatic Translation of Languages. In: Franz L. Alt (ed.) Advances in Computers, Volume 1, Academic Press (New York), pp. 91-163.

5. McCarthy, J. 1979. Ascribing Mental Qualities to Machines, Philosophical Perspectives in Artificial Intelligence, Harvester Press.

6. Chomsky, N. 1955. The Logical Structure of Linguistic Theory, distributed by Indiana University Linguistics Club.

7. Arnold, D.J., and L. Sadler. 1991. ìEurotra: An Assessment of the Current State of the ECís MT Programme,î Working Papers in Language Processing 32, Department of Language and Linguistics, University of Essex.

8. Vauquois, B. and C. Boitet. 1985. ìAutomated translation at Grenoble University,î Computational Linguistics, Vol. 11, 1, pp. 28-36.

9. Witkam, A.P.M. 1983. Distributed language translation: feasibility study of a multilingual facility for videotex information networks. Utrecht: BSO.

10. Landsbergen, J. 1987. ìIsomorphic grammars and their use in the ROSETTA translation system,î Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, Margaret King, ed., Edinburgh University Press, Edinburgh, pp. 351-372.

11. Schank, R.C., N. Goldman, C. Reiger and C.K. Riesbeck. 1973. MARGIE: memory, analysis, response generation and inferences in English Proceedings of IJCAI-73, International Joint Conference of Artificial Intelligence.

12. Kay, M. 1980. The proper place of men and machines in language translation. Research Report CSL-80-11. Xerox Palo Alto Research Center, Palo Alto, California.

13. Carbonell, J.G., R.E. Cullingford and A.G. Gershman. 1981. Steps towards knowledge-based machine translation IEEE transactions: PAMI, 3(4).

14. Nirenburg, S., V. Raskin and A. Tucker. 1986. ìOn Knowledge-Based Machine Translation',î COLING-86: Proceedings of the 11th International Conference on Computational Linguistics, Bonn, pp. 627-632.

15. Uchida, H. 1988. ìATLAS II: A Machine Translation System Using Conceptual Structure as an Interlingua,î Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Carnegie Mellon University, Centre for Machine Translation, Pittsburgh, Pennsylvania, USA, pp. 165-173.

16. Okumura, A., K. Muraki and S. Akamine. 1991. Multi-lingual sentence generation from the PIVOT interlingua, Proceedings of the MT Summit ë91, pp. 67-71.

17. Farwell, D., and Y. Wilks. 1991. Ultra: A Multi-lingual Machine Translator. Proceedings of the MT Summit III, 19-24.

18. Goodman, K. and S. Nirenburg. 1991. The KBMT Project: A case study in knowledge-based machine translation, Morgan Kaufmann Publishers.

19. Kaji, H. 1988. ìAn efficient execution method for rule-based machine translation,î COLING-88, Budapest, Vol. 2, pp. 824-829.

20. King, G.W. 1956. Stochastic Methods fo Mechanical Translation. Mechanical translation 3(2), 38-39.

21. Nagao, K. 1990. ìKnowledge-Based Structural Disambiguation,î COLING-90: Proceedings of the 13th International Conference on Computational Linguistics, Helsinki, Finland, Vol. 2, pp. 282-287.

22. Hutchins and Lovtsky, in press.

23. Hutchins, J. 1986. Machine Translation: Past, Present, Future, Ellis Horwood/Wiley, Chichester/New York.

24. Boitet, C. 1987. ìResearch and development on MT and related techniques at Grenoble University (GETA),î Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, Margaret King, ed., Edinburgh University Press, Edinburgh, pp. 133-153.

25. Melby, A.K. 1980a. A comparative look at junction grammar. In: The Sixth LACUS Forum (Columbia, S. Carolina: Hornbeam P.), 344-352.

26. Slocum, J. 1987. ìMETAL: the LRC machine translation system,î Machine Translation Today: The State of the Art, Proceedings of the Third Lugano Tutorial, Margaret King, ed., Edinburgh University Press University of Texas, Edinburgh, pp. 319-350.

27. King, M. 1982. ìEUROTRA: an attempt to achieve multilingual MT,î Practical Experience of Machine Translation, Veronica Lawson, ed., North-Holland, Amsterdam, pp. 139-147.

28. Pericliev, V. 1984. ìHandling syntactical ambiguity in machine translation,î COLING-84, pp. 521-524.

29. Kay, M. 1989. Head-driven parsing, Proceedings of the International Workshop on Parsing Technologies, June.

30. Ben-Ari, D., D.M. Berry and M. Rimon. 1988. ìTranslational Ambiguity Rephrased,î Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Carnegie Mellon University, Centre for Machine Translation, Pittsburgh, Pennsylvania, USA, pp. 175-183.

31. Isabelle, P. and L. Bourbeau. 1985. ìTAUM-AVIATION: its technical features and some experimental results,î Computational Linguistics, Vol. 11, 1, pp. 18-27.

32. Amano, S., H. Hirakawa, H. Nogami and A. Kumano. 1989. The Toshiba Machine Translation System Future Computing Systems, 2(3), pp. 227-246.

33. Nagao, M. 1989. Machine Translation, Oxford University Press.

34. Patel-Schneider, P.F. 1989. A four-valued Semantics for Terminological Reasoning, Artificial Intelligence, Vol. 38.S

35. Wilks, Y.A. 1990. Where am I Coming From: The Reversibility of Analysis and Generation in Natural Language Processing. In Martin Putz (ed) ìThirty Years of Linguistic Evolution,î Amsterdam: John Benjamins.

36. Brown, P. *et al.* 1990. ìA Statistical Approach to Machine Translation,î Computational Linguistics, Vol. 16, 2, pp. 79-85.

37. Jelinek, F. and R.L. Mercer. 1980. Interpolated Estimation of Markov Source Parameters from Sparse Data Proceedings of the Workshop on Pattern Recognition in Practice, May.