[From: *Annual Review of Information Science and Technology*, vol.19, 1984]

Machine Translation: A Contemporary View

ALLEN B. TUCKER, JR.
Colgate University

SERGEI NIRENBURG
Colgate University

INTRODUCTION

Because the early history of machine translation (MT), roughly from 1945 to 1970, is well documented elsewhere (HUTCHINS, 1978; 1982) and because several recent studies and advances in MT have eclipsed those past events and opinions, we begin this review with the contemporary state of the field. This chapter is divided roughly into two parts. In the first part, we discuss contemporary theoretical and practical issues surrounding MT. In the second, we evaluate 12 current MT research and development (R&D) efforts, including ongoing work in the United States, Europe, and Japan. Due to limitations of space we do not survey all such projects. We apologize for not devoting a separate section, for example, to SUSY and GETA, the two well-established European projects. Still, many of the ideas that originated in these projects are discussed in the section on transfer systems. Our discussions are also limited by our ability to obtain complete (and unclassified) documentation for some MT projects. The 12 projects that we cover in some detail are: Georgetown, SYSTRAN, SPANAM, TAUM, two projects from Kyoto University, LRC, EUROTRA, DLT, PHRAN-PHRED, MOPTRANS, and TRANSLATOR.

The authors thank the reviewers, especially Don Walker and Jonathan Slocum, for their insightful comments on earlier drafts of this chapter. Patricia Ryan deserves our appreciation as well for her extensive work on the manuscript (especially the bibliography). The authors, of course, are responsible for all attitudes and conclusions represented herein.

Annual Review of Information Science and Technology (ARIST], Volume 19, 1984
Martha E. Williams, Editor
Published for the American Society for Information Science (ASIS)
by Knowledge Industry Publications, Inc.

PRINCIPLES AND IMPLEMENTATION OF MACHINE TRANSLATION SYSTEMS

Here we discuss three theoretical and four implementation issues surrounding contemporary MT systems. The theoretical issues are the choice of MT strategy, grammars and parsing, and related computer science topics. The implementation issues are dictionary organization, software support, fields of discourse, and evaluating the quality of translations.

Machine Translation Strategies

Today the major strategic decision available to workers in MT is the choice between the transfer approach and the interlingual approach. Under the transfer approach a sentence in the source language (SL) is analyzed into an internal representation, after which a transfer is made at both the lexical and structural levels into corresponding structures in the target language (TL), and then the translation is generated. The transfer model is pictured in Figure 1. Three dictionaries are needed: 1) an SL dictionary, 2) a transfer dictionary, and 3) a TL dictionary. The approach is an improvement over the earlier "direct" translation systems, in which no structural information was used—i.e., the transfer was lexically driven and was not differentiated from the analysis or generation phases.

The level of transfer differs from system to system—the representation varies from purely syntactic deep structure markers to syntactico-semantic

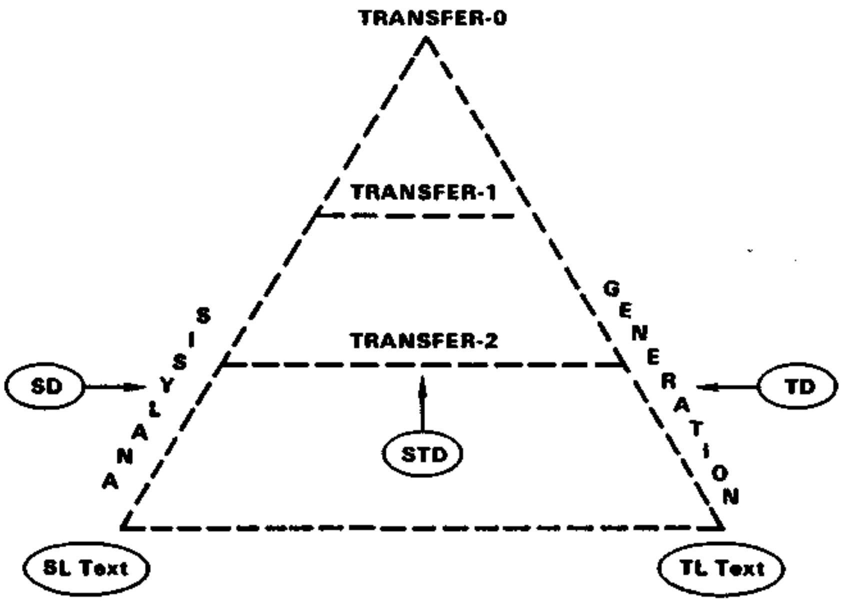

Figure 1. Variants of the transfer model of machine translation. SL and TL denote the source language and target language. SD, TD, and STD denote the source, target, and source-target dictionaries. This figure was adapted from VAUQUOIS.

(compositional semantics, case frame information, and so forth) annotated trees. Note that the transfer stage involves a (usually substantial) bilingual component—i.e., a component tailored for a specific SL-TL pair. This entails relative inefficiency in a multilingual environment since a transfer block will have to be written for *every* such pair. The tactical decisions necessitated by the transfer approach thus include: 1) choosing the level of transfer (cf. Figure 1), and thus 2) determining the relative "weight" of the monolingual vs. bilingual parts of the process—i.e., by trying to shift part of the work of the transfer block out to the SL analysis and TL synthesis stages, one can improve the cost-effectiveness of possible extensions to the multilingual aspect. In Figure 1 the extent and complexity of the transfer module range from TRANSFER-2 (maximum) through TRANSFER-0 (minimum). The transfer approach has been developed, maintained, and popularized by such well-known MT groups as GETA in Grenoble, France (BOITET; VAUQUOIS) and SUSY in Saarbruecken, Germany (MAAS).

An alternative to the transfer approach is to make the translation with the help of a universal, language-independent" representation of the text—an "interlingua." In effect, an interlingua permits the size of the transfer module in Figure 1 to be reduced to zero, and the MT model thus has two phases: analysis and generation. What makes this approach more attractive than transfer? First we can in principle dispense with bilinguality. Indeed, for a multilingual system with n SLs and m TLs the transfer approach will require mn (on the order of n^2) transfer blocks (if the sets of SLs and TLs are disjoint) in addition to n analyzers and m generators. In the interlingua approach, only n parsers and m generators will be needed (of a different sort, however).

One early suggestion in MT was to use an interlingua of a different kind: a natural language to serve as the pivot in a multilingual translation system. Thus, if Greek were chosen as the pivot for a seven-language multilingual design, the system would involve only 12 (on the order of n) transfer modules and 6 each of analysis and generation modules (see Figure 2).

As Figure 2 shows, however, the suggestion really implies 12 separate systems—i.e., each arrow implies a separate analysis, transfer, and generation phase. Therefore, we do not consider this a true interlingua approach. Curiously enough, some remnants of this type of thinking can be discerned in a recent MT proposal (see below).

The "true" interlingua projects fall into two classes: 1) the early syntactic approaches and 2) those inspired by artificial intelligence (AI). The former (e.g., the early efforts at the University of Texas and at Grenoble (LEHMANN & STACHOWITZ; VAUQUOIS) occurred chronologically before the advent of the transfer approach and provided valuable practical experience with the complexity of the interlingua concept. The main idea was very attractive: the (syntactic) structure obtained after parsing an SL text was declared universal (interlingual) and was supposed to be used directly by the generator. The bilingual dictionary for transferring lexical components remained intact in these systems, so that the separation of SL and TL was not complete. "In retrospect, the interlingual approach was perhaps too ambitious at that time (1960's)" (KNOWLES, p29), mostly because the expressive power of the syntactic representation was hardly sufficient to support the concept of an interlingua.

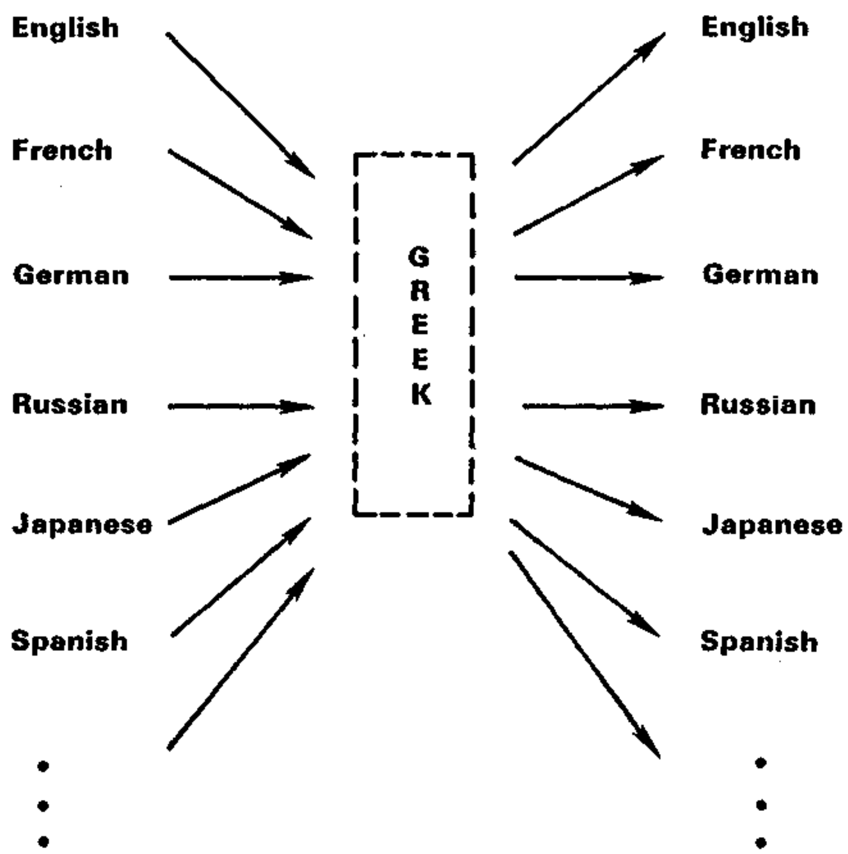

Figure 2. Interlingua model of machine translation, using natural language (arbitrarily Greek) as the interlingua.

A genuine interlingua must be able to express the *meaning* of the text to be translated. This type of research toward an interlingua-based MT system can proceed despite the debate about whether it is possible to capture and formalize the human encyclopedic knowledge that is a necessary part of language understanding. This is because one can work with "subworlds" and sublanguages in the hope of producing a translation scheme that can be extended beyond the experimental stage. The general scheme of the true interlingual approach is shown in Figure 3.

The methodology used in this type of project is borrowed almost exclusively from AI, for which MT could be (but curiously enough has not yet become) a major application. It is quite understandable historically that in the 1960s, concepts on which to base MT were borrowed from linguistics, which was then in vogue and showing much promise. Today AI appears, at least to nonspecialists, to be a much more glamorous field than linguistics. Workers in applications such as MT, in which knowledge about language processing is important, more often look to AI for the solutions to their problems. Therefore, it is not surprising that, for example, SAWAI ET AL. present their Japanese-English MT system, ATLAS/I, as featuring "knowledge representation," even though their linguistic analysis is borrowed from JACKENDOFF and their methodology is basically transfer. Another example of an experimental interlingua system is the program by CORDIER & MOGHRABI, which translates cooking recipes from French into Arabic. The authors use semantic primitives to represent the meaning of the input text. There is no clear statement about the organization of these primitives. At any rate, in the specialized subworld (and sublanguage) chosen for the application, the number of actions around which the representation is built is small enough to be enumerated, so that a representation scheme with less structure can be used. Such approaches can be envisaged for other small subworlds and sublanguages (e.g., meteorological reports).

There has been some interest in MT as an application within the AI community (CARBONELL ET AL.; LYTINEN & SCHANK; WILENSKY & MORGAN). The main strategy of such efforts is to equate utterances in an interlingua with formulas of a knowledge representation scheme (in practice, a version of the conceptual dependency representation language augmented with higher-level structures—e.g., scripts, plans, goals, memory organization packets (MOPs)). The process of translation is assumed to proceed along the following lines. "First, the source text is analyzed and mapped into a language-free conceptual representation. Inference mechanisms then apply contextual world knowledge to augment the representation in various ways, adding information about items that were only implicit in the text. Finally, a natural-language generator maps appropriate sections of the language-free representation into the target language" (CARBONELL ET AL., p376).

CARBONELL ET AL. say that their computer programs "technically speaking. . .do not perform strict translation, but rather retell or summarize the source text in the target language" (p377). It is clearly necessary to provide a theory of what sections are appropriate or what to retell and what to include in a summary and what to leave out—a theory of "salient features" of the text. The AI practitioners' interest in MT is a very positive development, even though "practical MT has not been a primary working goal with.. .the .. .understanding systems built within the conceptual dependency/knowl-

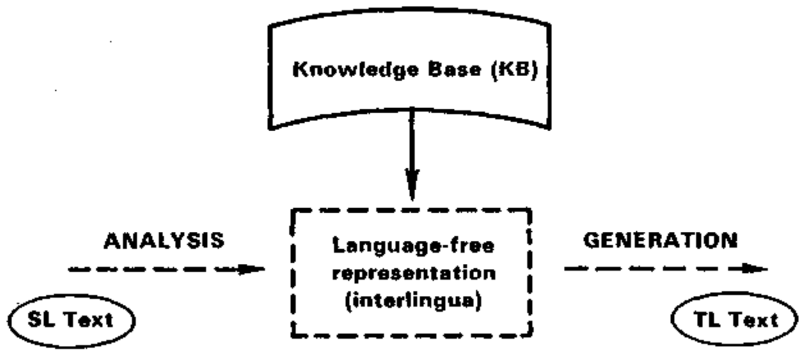

Figure 3. The "true" interlingua model of machine translation.

edge structure framework" (CARBONELL ET AL., p376). AI-based MT efforts are discussed in more detail in the section on PHRAN-PHRED and MOPTRANS.

An important methodological question is whether the particular knowledge representation scheme used can be applied directly to MT or whether it must be customized. For example, it is reasonable to investigate ways to augment a typical AI knowledge representation scheme with explicitly presented linguistic (syntactico-semantic) knowledge as well as a representation of the expert behavior of human translators so that it can become effective in the MT domain. In our opinion, the general techniques of knowledge representation are not immediately applicable to natural language translation. The customization required by MT demands direct research on its merits.

Grammars and Parsing

It was recognized early in MT that a systematic analysis of the input text is an indispensable part of the translation process. Morphological analysis—the determination of word class and word form to which an input word belongs—has since become a theoretically less controversial task. The task of syntactic analysis—identification of constituent and/or dependency structure of input text sentences—has not yet been agreed on; different approaches coexist, and none has emerged as the definitive method.

Parsing natural language was probably the most important and one of the most widespread topics of research in AI from roughly 1968 to 1978. WINOGRAD (Chapter 7) gives an excellent review of (predominantly syntactic) parsing systems. SPARCK JONES & WILKS edited a recent collection devoted exclusively to parsing. The most widely used grammar formalisms on which the various parsing systems are based include augmented phrase structure grammars, transformational grammars, active chart formalisms, augmented transition networks, and "situation-action rule" systems that do not use a grammar in the traditional linguistic and computational sense.

The form of the structures assigned to sentences can also vary. The most widely used structures are: various annotated surface structures (either immediate constituent, or dependency, or a mixture of the two); variants of the deep syntactic structure in the sense of transformational grammar or case grammar; and various systematic nonsyntactic structures; such as conceptual dependency representations produced by ELI (e.g., SCHANK & RIESBECK) or Wilks's formulas (WILKS, 1975).

At present most of the MT systems parse SL text into a structure that provides: 1) sentence constituent information, and 2) case frame information for verbs and nouns. In the terms of CHARNIAK (1983) these processes belong to a parallel (intermingled) application of syntactic and "non-inferential-semantic" rules. An example of the usual "syntax plus case frame" approach is the LRC parser (SLOCUM, 1982).

If an MT system strives to obtain a true interlingua representation, the process must be augmented by "compositional semantics," the rules of which permit the construction of the "logical form" of the input, as suggested by CHARNIAK (1983). Only at this stage, when the process of parsing ends, can the full advantages of knowledge representation (interlingua) be fully felt.

With the help of the rules of "inferential semantics" one may augment, through a chain of knowledge-based deductions (inferences), the system's understanding of the meaning of the input and thus provide a high level of reliability to the generator. Examples of "deeper" parsers are the members of the ELI family (BIRNBAUM & SELFRIDGE) that produce conceptual dependency (one type of "logical form") structures. As of now there is no reasonably scaled MT system that uses knowledge representation for inference making in this way. Despite their use of a frame-based approach, SAWAI ET AL. actually implement a transfer system and claim no inferencing capability. JOHNSON gives a thoughtful discussion of the properties necessary for an MT parser and of design choices.

The computational peculiarities of parsers are seldom addressed in reports about MT systems. A notable exception is the LRC project, in which specific attention was paid to the choice of the parser. "The current METAL parser is a variation on the Cocke-Kasami-Younger bottom-up algorithm.. .augmented with top-down filtering.... This parser was shown to be highly efficient during an extensive series of experiments comparing a dozen parsers on the basis of their practical performance characteristics" (SLOCUM & BENNETT).

Computer Science and Machine Translation

In addition to AI, three traditional areas of computer science have direct bearing on MT research: 1) database technology; 2) programming languages, and 3) computer architecture. For robust MT systems to be developed, it will be necessary to effectively direct current database technology to the task of dictionary design and information retrieval. Because of the limited extent of their models, present AI experiments do not effectively address these areas. One of the weaknesses of LISP (the programming language most often used in AI research) is its lack of file processing or database management facilities. In LISP, one assumes that all s-expressions (both programs and data) are in "virtual" memory, which is theoretically limitless. Garbage collection becomes an important and time-consuming component of AI systems (for a report of practical experience see SLOCUM & BENNETT). The issue of using secondary storage for large AI databases has not been adequately addressed. The state of the art in merging the database theory and practice with AI is reflected in the following statement: "If your [AI] database reaches a size where secondary storage is needed, then you have entered *terra incognita;* no one knows how efficient a large, random-access database can be. On the other hand, few people are willing to bet that making such a database efficient is impossible" (CHARNIAK ET AL., p220).

Current operational MT systems (such as SPANAM) are implemented using computer software that supports so-called "indexed" files. Here, the dictionary resides on direct-access secondary storage, and all entries can be immediately accessed, using lexical information (the word stem) as the key. This is known as content-addressable memory and is a minimum requirement for effectively implementing large random-access dictionaries for MT.

In a more sophisticated system, the theoretical tools and flexibility offered by relational database systems and query languages would seem to be appropriate. LISP machines offer database software (the FLAVORS package can be used for this purpose) as a companion tool for building AI knowledge bases for intelligent systems. Current and projected MT systems have yet to effectively integrate such technology for dictionary maintenance and access.

Current operational MT systems are written either in assembly language or in PL/I. Most experimental systems, on the other hand, are implemented in LISP. To develop MT systems effectively in the future, better programming-language support needs to be identified—i.e., a language that allows effective text processing, dynamic data structure management, and random-access dictionary information retrieval is necessary. PL/I has all of these attributes, but its relative inefficiency and unwieldiness make it unsuitable for most programmers. Recent advances in the design of programming languages, including C, ADA, PROLOG, and newer implementations of LISP, may provide more effective language tools for MT than the traditional languages have done.

Finally, effective MT systems must be implemented on computer hardware that not only supports such languages and database technology but also has storage capacities and the character-set flexibility to accommodate large knowledge bases (dictionaries) and diverse language typographies. Traditional MT systems (e.g., SYSTRAN and SPANAM) are implemented on IBM computers, running in batch mode and using attached word processors and optical character readers (OCRs) to facilitate text input, post-editing, and document preparation. An exception, the Weidner System (HUNDT), runs interactively on a minicomputer especially adapted for MT.

Recent developments in microcomputers and word processors make us optimistic about the availability of hardware that will be truly suitable for MT. Xerox Corp., for instance, manufactures word processors that support multiple alphabets, including Japanese, Arabic, and Russian as well as the European languages. This is an essential capability for any MT system that aspires to translate multiple languages with diverse alphabets. In the past, the Georgetown Russian-English MT system (ZARECHNAK) required the input text to be encoded from the Cyrillic to the English alphabet before submission to the translation process. This encoding is expensive, time consuming, and prone to error, and it is unacceptable as a long-term solution for production MT.

Some projects (SHIAO-SHU; SOMERS) are also looking at the possibility of adapting MT to the personal microcomputer, such as an IBM PC or a DEC (Digital Equipment Corp.) Rainbow. Personal computers now possess powerful software and adequate online storage to be seriously considered for MT. At this writing, about $7,000 buys a microcomputer with 512 K bytes of memory, 10 megabytes of online disc storage, a complete C or LISP programming system, and database management software that could support MT. Because diverse alphabets are supported, the necessary word processing can be done on the same machine. Finally, these machines can run in a stand-alone mode or can be networked in various ways, so that both high- and low-volume translation can be accommodated. The idea that MT can be done on a stand-alone machine at any location in the world that has a simple electrical power outlet is thus a reality in terms of the hardware.

Dictionary Structure

The organization and content of the dictionaries for any MT system are determined mainly by the goals of the system, its linguistic strategies, and the limits of the computer system on which it resides. Current operational systems translate only a single language pair and thus must accommodate only the morpho-syntactic, semantic, and pragmatic information found in those particular languages.

For example, an MT dictionary for Spanish requires that each entry contain (at least): part of speech, gender, number, person, and tense. An English dictionary entry, however, might not contain gender since gender does not play a major role in English syntactic analysis or generation (it is used for pronominal reference determination). A Russian dictionary entry will include, for example, aspect information for verbs.

However, traditional MT dictionaries contain only as much information as is needed to effectively analyze source text and to synthesize target text. That includes the basic syntactic information noted above, together with some keys for dealing with prepositions, irregular forms, semantic case discrimination, and idioms. A SPANAM dictionary entry, which may be considered typical, contains 160 bytes, and there are approximately 60,000 such entries in the entire dictionary.

MT dictionaries usually store entries in stem form and leave it to the algorithm to automatically generate morphological variants. For instance, a Spanish or Russian adjective would be stored only in its masculine singular form; its other three forms would be derived by morphological routines. A regular English verb, say "push," would be stored only once, and the other forms (pushes, pushed, pushing) would be derived automatically. This kind of strategy is essential to avoid the potential explosion in dictionary size.

Another issue in dictionary design is whether or not to use separate source and target dictionaries or a single dictionary that contains, for each source language entry, all of its potential target language translations. Moreover, some systems use one dictionary structure for idioms, another for high-frequency words, and another for the rest. The choice is generally made on the basis of efficiency and programming convenience, and none is conceptually more effective than another.

A far more important issue is the matter of dictionary design for multilingual (e.g., EUROTRA) and knowledge-based MT systems of the future. Here the dictionary entry must be flexible, so that it can accommodate the semantic and pragmatic information for any of its various languages. Moreover, a common meta-language must be available so that dictionary encoders and linguists in the various languages have a mutually consistent means of communicating information about linguistic phenomena among diverse languages. The best approach methodologically seems to be to adopt or modify one of the existing knowledge representation languages as the meta-language in which to formulate dictionary information. Unfortunately, this has not been a primary interest either among AI researchers or MT developers. One exception is the philosophy of TRANSLATOR (see below).

Researchers in AI have discovered that the implementation of truly useful knowledge-based systems, which purport to have some reasonable level of "understanding" of the text they are "reading," requires immense quantities of knowledge. The information storage and retrieval problems that accompany such comprehensive knowledge representation are not well understood. It appears, however, that this kind of "intelligent" dictionary and related data structures will expand the storage requirements of conventional MT systems by a significant factor. Current MT systems for a single language pair maintain dictionaries whose size is on the order of 10 megabytes of online storage.

Supporting Software

A prerequisite for the effective implementation and refinement of an MT system is the availability of effective supporting software to maintain dictionary information and linguistic rules. A common mistake has been to underestimate these requirements and to ignore the fact that once an MT system is installed, the user is more apt to make significant refinements to it than the developer. One of the major reasons for the recent success, measured intuitively, of the SYSTRAN and SPANAM systems is that the users have complete and convenient mechanical access to the dictionaries and linguistic routines (PIGOTT; VASCONCELLOS) and constantly refine the systems as they gain experience.

Many of the tools for updating dictionaries are available in existing database management software. Others, such as the ability to modify grammatical parsing rules, are not. The SYSTRAN system, however, has a facility that allows a segment of assembly language instructions to be translated into a coherent English-language "linguistic statement" of the algorithm being performed. This statement is much more intelligible to the linguist than is the corresponding assembly language code. Once the linguistic statement is appropriately modified by the linguist, the corresponding assembly language modifications are made in the program itself. A comparable strategy is implemented experimentally in the LRC project (SLOCUM & BENNETT).

Even the developers of new systems (NAGAO, 1982) report difficulty in training dictionary workers to properly and efficiently encode semantic case information into dictionary entries. Much experience seems to be needed before the quantum jump is made to build effective dictionaries that can be properly called "knowledge bases" for machine translation. Proper integration of linguistic information with world knowledge and common-sense inferencing schemes seems to be a highly technical skill that cannot be taught overnight to those who are responsible for building and refining these knowledge bases.

Fields of Discourse

In general, MT systems limit their domain of usefulness to a particular field or area of discourse. For example, the TAUM-METEO system (ISABELLE) is applied only to weather reports and thus has a limited vocabulary. The Georgetown system (ZARECHNAK) concentrates on nuclear physics, and the SPANAM (TUCKER; VASCONCELLOS) system focuses on public health. Current developers are also aiming their MT systems at well-defined and highly stylized text, such as auto repair manuals and computer science publications.

Such a focus generally gives the MT system a stronger chance for success because the development of a highly specialized and complete dictionary in a specific field is needed to compensate effectively for various inabilities of the system to really "understand" the text itself. Such text, in other words, tends to be self-clarifying since it has a relatively high frequency of technical terms, each of which renders only a single interpretation, and a relatively low frequency of ambiguous terms and difficult style.

It remains to be seen how much the introduction of "understanding" to MT can widen the translation domain. Even the recent AI experiments (CARBONELL ET AL.) are not particularly ambitious in this respect—they stick to examples about visiting restaurants and reports about automobile accidents.

Performance and Evaluation

The performance of operational MT systems is usually measured in terms of their cost per 1,000 words and their speed in pages per post-editor per hour vs. the relative cost and speed of human translation. Some specific information on cost and speed is reported in LAWSON, and some more recent information is given in the next section.

In our opinion, it is becoming increasingly uninformative to compare the performance of MT systems with that of human translators, even though large organizations must do just that to justify their MT investments. In the long run, machine translation systems and human translators may not be viewed as competitors, with humans concentrating on "artistic" translation, simultaneous interpretation, and so forth.

More important is the question of translation quality—i.e., the fidelity of the translation to the original text and the legibility of the translation. Much discussion of quality also appears in LAWSON and elsewhere, but it seems that no effective, universally applicable, quasi-objective measure of translation quality—either human or machine generated—has yet been discovered.

The need for such a measure will be increasingly important in the future, not for comparing machines with humans but for comparing various types of MT systems with each other to determine the relative effectiveness of different translation strategies. A proposal for developing such a measure has been suggested by NIRENBURG & TUCKER.

SELECTED RESEARCH, DEVELOPMENT, AND OPERATIONAL MT SYSTEMS

Here we summarize the characteristics of several current and recent MT systems. Some of these systems are operational (SYSTRAN, SPANAM, and TAUM-METEO); others are in various stages of development (LRC, EUROTRA, and others); still others are properly categorized as research models (see LYTINEN & SCHANK; NAGAO ET AL.; TUCKER & NIRENBURG; WILENSKY & ARENS).

Georgetown, SYSTRAN, and SPANAM Systems

The Georgetown MT system (ZARECHNAK) is the first truly successful effort to develop an operational MT system. Its approach, labeled as "first generation" or "direct translation," was later emulated by the SYSTRAN system (PIGOTT; TOMA) and the more recent SPANAM system (TUCKER;

VASCONCELLOS). The Georgetown system went into production, translating Russian to English, in 1964 at Oak Ridge National Laboratory in Tennessee. The SYSTRAN system has been in use for Russian-English translation at Wright-Patterson Air Force Base since 1970 and at the European Economic Community (EEC) Headquarters in Luxembourg since 1976. SYSTRAN is used to translate English into French, Italian, and German and to translate French and German into English. The SPANAM system translates Spanish into English, and has been in operation since 1980.

The linguistic methodology in these systems is portrayed below. Here the text is taken one sentence at a time and passed through several steps:

1. Dictionary lookup and morphological analysis,
2. Homographs,
3. Compound nouns,
4. Phrases,
5. Idioms,
6. Prepositions,
7. Subject-predicate identification,
8. Ambiguities,
9. TL synthesis, and
10. Rearrangement.

Actually, each system differs slightly in the order and use of these steps, but the general approach is the same for all of them. For instance, the SPANAM system has separate source and target dictionaries, so that the "target synthesis" step includes target dictionary lookup. Moreover, some versions of SYSTRAN include the ability to use semantic case information to assist in subject-predicate identification and other disambiguation tasks. For more discussion of the details of these steps, readers should see TOMA, TUCKER, and ZARECHNAK.

The effectiveness of these systems relies principally on three factors: 1) highly developed dictionaries and morphological analysis routines, 2) human post-editing of the raw translation before distribution of the results, and 3) well-developed word- and text-processing tools to aid the post-editor and the dictionary officer. For instance, in the SPANAM system each Spanish noun, verb, adjective, and idiom is stored in the dictionary in its stem form; all other forms are dynamically derived by morphological analysis routines.

Although operational MT systems rely on post-editors to "clean up" the raw translation, they do not rely on any pre-editing of the text. Moreover, since the word-processing activities are directly connected to the main computer that performs the machine translation, any word-processed document is an immediate candidate for MT without additional preparation cost.

The productivity and volume of output for operational MT systems vary widely. The Georgetown system, oldest of the three, has produced virtually hundreds of thousands of (250-word) pages of Russian-English translations since 1970. The SPANAM system, in use since 1980, has produced over 5,000 pages of Spanish-English translation. SYSTRAN, since 1980, also has turned out over 5,000 pages of translation among its several language pairs. The productivity of a post-editor has also been measured in these latter two systems:

for SPANAM it is estimated at 6,000 words per day (VASCONCELLOS), or three pages per hour. For SYSTRAN it is estimated at two pages per hour (PIGOTT). Actual cost per page of post-edited MT is difficult to estimate, but both organizations cite significant cost and time savings over manual translation.

TAUM

The MT project at the University of Montreal, called TAUM, was active between 1968-1980 and yielded a number of experimental systems (TAUM-71, TAUM-73, TAUM-76, and TAUM-AVIATION) and one operational system (TAUM-METEO). TAUM-METEO may be today the closest approximation of a fully automated high-quality translation system among those that are operational. The methodology of the TAUM project as a whole is transfer. The transfer component involves two subcomponents: lexical and structural. A number of formalisms and programming environments have been developed: Q-SYSTEMS (COLMERAUER) facilitates the linguist's work on specifying the grammar rules; SISIF (MORIN) aids pre- and post-processing of the text; REZO (STEWART) is a modification of the augmented transition network grammar system for syntactic analysis. Analysis in TAUM methodology is syntactic, without lexical decomposition, but it involves a number of semantic features (e.g., subject-verb agreement, and semantic features on verbs). Synthesis is performed on a sentential basis, so that the problems of focus and anaphora remain unsolved.

The major practical success of TAUM was in the TAUM-METEO application. Faced with the necessity of developing an operational system, the members of the TAUM group chose in 1974 to build a system to translate weather reports issued by the Canadian Environment Department from English into French. "The feasibility studies, design, development and on-site implementation of an operational version of the system took less than two years (approximately 8 person/years)" (ISABELLE, p27). The most striking peculiarities of TAUM-METEO are its overt semanticity and the lack of the transfer module.

Analysis in TAUM-METEO is based on a semantic grammar (cf. BATES; see BURTON for discussion), in which the nonterminal vocabulary includes not only the familiar set of syntactic labels, such as NP, VP, and so forth, but also domain-specific semantic markers. One such marker can be "atmospheric_ condition," which "consists of a weather condition optionally modified by a locative or temporal specification; but the condition itself cuts across syntactic categories: 1) MAINLY SUNNY TODAY; 2) A FEW SHOWERS THIS EVENING" (ISABELLE, p29). A semantic grammar subtree corresponding to one of the possible derivations from the node "atmospheric_condition" is shown in Figure 4.

Transfer in TAUM-METEO is effectively incorporated into analysis. The few operations that are not covered in analysis (e.g., the correct placement of French adjectives) are dealt with in the synthesis stage, which is essentially trivial because of the peculiarities of the sublanguage of translation.

Unlike any other TAUM efforts, or, for that matter, any other transfer system, TAUM-METEO was designed to operate from the outset in an extremely narrow sublanguage (1,500 dictionary entries, including several hundred place

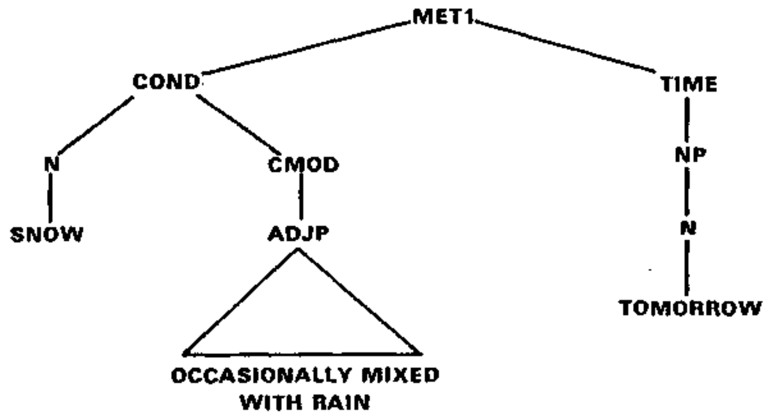

Figure 4. A semantic grammar tree from TAUM-METEO

names; input texts containing no tensed verbs). It therefore uses unabashedly ad hoc measures to improve efficiency. The decisions to discard the transfer and to use a semantic grammar were made with efficiency in mind. The same applies to the decision to disregard morphological analysis (indeed, in so small a dictionary the space efficiency gained with the help of morphological analysis is negligible).

TAUM-METEO has been operational since 1977, translating about five million words annually, 80% of which do not need post-editing. Many of the remaining errors are caused by noise in the communication lines and misspellings.

The study of sublanguages and subworlds seems to us to be a most important theoretical topic for MT. The success and the experience of TAUM-METEO in this respect should attract the attention of other MT groups (see KITTREDGE & LEHRBERGER for a discussion of the sublanguage issue). In our opinion, it will require only a relatively small shift of emphasis to present the semantic grammar of the sort used in TAUM-METEO in terms of a knowledge representation scheme and operations on it. The result of input-text analysis will be an expression in this representation language; synthesis will take this representation as input and produce a TL sentence. It is not an unreasonable task to build a representation based on semantic primitives for 1,500 objects. Therefore, TAUM-METEO is potentially a precursor of sublanguage-oriented interlingual MT systems.

The Kyoto and ATLAS/I Projects

It is well known that the Japanese Fifth Generation project places strong emphasis on natural language translation as a major technical goal (MCCORDUCK). Within this framework, three substantial MT projects have gained significant momentum, with strong support from the Japanese government and industry.

One project, headed by Makoto Nagao at Kyoto University (NAGAO, 1983; NAGAO ET AL.), aims at a Japanese-English system to be developed within the period February 1982 to February 1985. The system will translate computer manuals and scientific abstracts from Japanese into English. It is a transfer system, but each of the stages—analysis, transfer, and synthesis--has conservative first-generation components. The system is based on the premise that most linguistic phenomena can be handled by lexical rules rather than syntactic rules. The developer feels that too much attention has been paid in other systems to the incorporation of pragmatic knowledge.

Thus, the linguistic method depends strongly on the lexicon. The analysis phase translates source text into "dependency structures," based on case grammar, via four steps: 1) morphological analysis, 2) segmentation, 3) fragment relationships, and 4) noun-phrase analysis. This analysis phase finally produces Japanese Intermediate Structure (JIS), which is a normalized intermediate form. The transfer phase then takes JIS into EIS (English Intermediate Structure), a form susceptible to synthesis via a conventional phrase structure grammar. Three dictionaries are thus needed: one for Japanese analysis, one for Japanese-English transfer, and one for English synthesis.

The system is implemented in LISP (although PL/I is used for the morphological routines) on a FACOM M-200 computer. The software contains powerful components for manipulating tree structures (transfer phase) and rewriting rules (synthesis phase). It also has a utility called GRADE, which is a tool for compiling linguistic/grammar rules into internal format for processing. No experience has been reported for this system, but its ambitious deadlines should force some results soon.

The second Japanese project, also at Kyoto University, is headed by T. Nishida and S. Doshita (NISHIDA; NISHIDA & DOSHITA, 1982; 1983). It aims to translate computer manuals from English to Japanese and is also based on the transfer approach. However, the stages of analysis, transfer, and synthesis are quite different in their content from the first project. In the analysis phase, a version of Montague grammar (MONTAGUE) is used to translate English text into so-called EFR (English Formal Representation), which is a kind of "normal form." For example, the sentence, HE COMES LATE, would translate into: (HE (LATE(COMES))).

Next, the transfer phase takes EFR into the so-called Conceptual Phrase Structure (CPS) representation, a network representing the various fragments in the parse. CPS is a frame-based data structure in which the syntactic and semantic information of each Japanese lexical unit or phrase in the sentence is packed. For instance, the CPS representation of the sentence above would have the following elements:

```
NP        ADV        S
 :         :         :
HE        LATE      NP  VP
                     :    \
                     X   COMES
```

Finally, the generation phase recombines these fragments in ways appropriate to Japanese syntactic structure, using a collection of heuristic rules called REFORM.

The system is implemented in prototype form using LISP, and it has been used to translate short sample texts (40 sentences) into Japanese. Human assistance was used in the analysis phase to help resolve ambiguities in the English text. An excellent discussion of the system and display of its output are given by NISHIDA.

The third Japanese MT effort, called ATLAS/I, aims to translate software reports from Japanese into English (SAWAI ET AL.). It, too, uses a transfer approach but claims to utilize a frame-based knowledge representation scheme (called FKR-0) throughout the analysis, transfer, and generation phases of the translation process.

The analysis phase includes a preprocessing step, a phrase dictionary look-up, and a word dictionary lookup. This is followed by a routine for handling proper nouns, a syntactic and surface case analysis, a deep case analysis and structural translation, and finally an English synthesis step.

The use of a frame representation throughout the translation process appears to be a unifying and novel approach to MT. However, the depth and complexity of linguistic information embodied in this system are not clear from the information provided. Among the three Japanese MT efforts reported here, this one seems to be in the earliest stage of development. Thus, all three Japanese MT efforts are committed to the Japanese-English language pair and to the transfer approach in one way or another. None is sufficiently developed to be evaluated constructively.

The LRC Project

The MT project at the University of Texas at Austin has been active, in one form or another, since 1961. The current effort is the first attempt at a large-scale practical implementation. The translation is from German to English in the field of telecommunications. The LRC project is supported by Siemens Corp. The implementation "is nearing the status of a production system (a version will be delivered to the project sponsor soon after [December 1982])" (SLOCUM, 1982, p1). The LRC approach uses the transfer method, with monolingual source and target dictionaries and a bilingual transfer dictionary and is implemented in LISP.

Transfer dictionaries consist essentially of "canonical word pairs" connecting the stems in SL and TL, "augmented by an arbitrary collection of context restrictions" (SLOCUM & BENNETT, p10). The context restriction can be of both a syntactic and semantic nature (e.g., nouns belong to certain semantic types). There is no discussion of the relative importance and size of the monolingual and bilingual processing although the transfer stage is not as "minimized" as is planned for EUROTRA (KING, 1982).

The grammar used is a version of phrase structure grammar with a transformational component. Case frame constraints are used as well. The result of parsing is a syntactic structure tree with some semantic information attached (e.g., class membership of lexemes, case information). Grammar rules are augmented with texts and the CONSTR part, which makes them similar to ATN arcs. One structural difference from ATN arcs is the TRANSFer part of a METAL grammar rule, which is used only after a sentence has been parsed, at which time "the system will perform the operations specified, generally moving down the tree to the terminal nodes where lexical substitution takes place" (SLOCUM & BENNETT, p14).

Transformations can be applied to any of the three parts of the grammar rules. The transformations "range from simple movement and deletion operations to highly complex transformations that add structure, perform tests, etc." (SLOCUM & BENNETT, p14). This is the place where an indefinite amount of ad hoc knowledge can be introduced into the system. Neither the conditions nor the mechanism of applying transformations is discussed (as suggested by WINOGRAD, Chapter 4).

The fact that syntax and semantics are not formally demarcated in parsing is a distinguishing point of LRC. Computationally the parser is (as of July 1982) a version of the Cocke-Kasami-Younger bottom-up algorithm augmented with top-down filtering. In other words, it is a version of an active chart-parsing scheme (KAPLAN). The parser was actually chosen over a dozen alternative parsers, a chore not many MT workers have undertaken.

The parser operation depends strongly on the preliminary lexical and morphological analysis. It generates in-process definitions (presumably, slot-filler type) for nonwords and unknown words and treats parentheticals through a recursive call to the parser itself (a clever idea).

A special routine within the parser rates the variant parses as to their applicability. This is another point at which the knowledge of the world and the expert knowledge of the type possessed by translators about their craft can be subsequently added to the system.

A most important feature of the parser is that it is robust; it will attempt to parse and eventually to translate a sentence that contains an unknown word or an ungrammatical sentence. This feature seems to be the main reason for choosing a bottom-up parser.

In the LRC reports much less space is devoted to the transfer stage of the translation process than to the parsing stage. It appears that the rules originally used for recognizing a constituent are retained until the transfer stage, at which time they are used again to help the transfer process. Unlike parsing, transfer proceeds in a top-down fashion because no failures are anticipated.

In the LRC reports the generation stage is given even less space than the transfer stage. The generator appears to operate by "simply taking the TL tree…and appending together all of the lexical allomorphs...located in its terminal nodes" (SLOCUM & BENNETT, p17).

AI methodology is present in an oblique sense at best. The LRC system does not seem to use any significant AI methodology in a functional way; the algorithm exhibits no level of "understanding" in the sense that it cannot add implicit information to what is said, nor can it answer questions about the content of the text. Moreover, no attempt is made by the system to explicitly emulate the expert behavior of the translator beyond the morphological and syntactic levels.

As of July 1982, experiments were performed on translating 43,000 words of German technical manuals into English. The DEC 2060 implementation (which translated "330 pages of texts, in three segments, over a two-year period") "was measured at slightly under 2 seconds of CPU time per input word" (SLOCUM, 1982, p10), 45% of which is typically spent on garbage collection by the LISP system. The LM-2 LISP Machine implementation is expected to be initially one-fifth to one-sixth as fast as the DEC, but with the new generation of LISP machines and additional memory, this rate is expected to improve.

Recognizing that the standards for measuring machine translation quality are little understood and vary widely, the LRC report concedes that the "qualitative results are all to be measured by professional technical translators employed by the project sponsor" (SLOCUM, 1982, p11).

In terms of the amount and quality of support tools for the translating program, the working environment of the LRC system is probably the most complete and best designed among the existing MT systems, both production and experimental ones. Adequate attention has been paid to auxiliary tasks, such as optimizing the process of compiling dictionaries using the advances in database theory and practice, introducing a spelling-correction module, benchmark tools, and text-processing support. The attitude toward the translation program as the innermost routine in a complex software system seems to be a fruitful approach from the software engineering standpoint. It is also one of the salient features of LRC. (For more details, see SLOCUM, 1982. p9). The LRC system design reflects a meticulous attention to detail, modularity, efficiency, and human engineering. In this respect the approach cannot be praised enough.

Much less attention is paid by the developers to theoretical and methodological principles. In their discussion Slocum and Bennett unnecessarily defend the choice of a bottom-up parser, a phrase structure grammar, and feature-based semantics. These choices are not truly iconoclastic. To us, the only important point that is not adequately defensible (in terms other than deadlines) is the overt transfer approach. Unfortunately, this is the central strategic choice in the design of any MT system. Even from the economic point of view the transfer approach has never been demonstrated to be effective in any but a severely restricted domain. Multilinguality will be even more expensive, as the experience with the EEC EUROTRA project will doubtless show. The transfer module in LRC is quite deep, as shown by the involvement of all the grammar rules used in parsing at the transfer stage.

In defense of transfer, the LRC progress report cites that "no adequate description of universal deep structure has been proposed" (SLOCUM, 1982, p19). However, by the same token, no adequate grammar for natural language has been proposed either. In light of the recent developments in linguistics, At, and software technology, we are convinced that MT system designers can produce an interlingua-based MT system with a less-than-perfect interlingua and still deliver a better result than a transfer-based system can. Most importantly, any serious attempt at an interlingua-based MT system will fundamentally advance the state of the art and provide a significant step toward an optimum solution for the general problem of machine translation.

EUROTRA

EEC has recently begun to support a new and ambitious concept in machine translation. Dubbed EUROTRA (KING, 1981; 1982), it is a project to develop a multilingual MT system that will translate any of the seven official EEC languages into any other. These languages are English, French, German, Dutch, Danish, Italian, and Greek. The project involves linguists in most of the cooperating countries, and a working system is expected by the early 1990s.

The framework for translation is proposed as a generalization of the basic transfer approach, as described earlier. However, the multilingual capability for seven languages requires seven analysis modules, $7 \times 6 = 42$ transfer modules, and seven generation modules, as shown in Figure 5.

The rationale for this approach is both practical and theoretical. On the practical side, the effort required to develop the 56 modules for seven languages will be significantly less than that which would be needed to develop a separate MT algorithm for each of the 42 different language pairs. This reasoning also assumes that the size and complexity of the 42 transfer modules will be minimal. On the theoretical side, this model permits, even encourages, separation of methodology among the different language working groups. Thus, for instance, one language may be analyzed using a phrase-structure grammar while another may use ATNs for the same purpose. Considering the diversity of current opinion on parsing natural language (cf., e.g., SPARCK JONES & WILKS), this kind of flexibility is very important.

To date, EUROTRA has not been described in technical detail in unclassified literature. This project appears to be a difficult task, not only from the theoretical/implementation point of view but also from the organizational/management point of view. If successful, EUROTRA will certainly set a new standard for comparison in machine translation and will distinguish itself from first-generation predecessors in many ways.

DLT

Distributed Language Translation (DLT) is a project being developed in Utrecht, The Netherlands. Its developers have just completed a feasibility study (WITKAM). They aim to develop multilingual MT using an interlingua model. A pilot study is planned to take two years (and 12 man-years), and the subsequent development stage is to take three more years. The availability of the product is planned for the early 1990s.

The pilot study expects to include work on a system of interlingua-to-German translation, with only a simulation of a source language parser

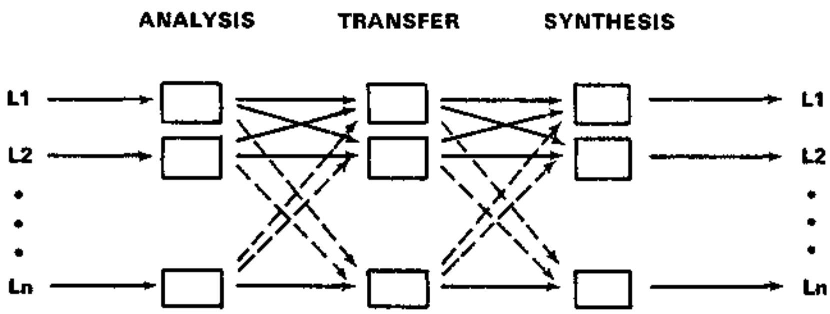

Figure 5. Multilingual transfer model of machine translation, as in EUROTRA

(through a dialog system). The pilot project will involve: 1) the specification of the intermediate language kernel (interlingua or IL) which involves writing an IL grammar in the ATN form, devising a parse tree structure, compiling an IL monolingual dictionary, testing, and so forth; 2) the development of the TL part, including the interlingua-to-German transfer dictionary, the German synthesis dictionary, etc.—borrowed from SUSY (MAAS); and 3) the accumulation of a compendium of terminology on international business and law (to enrich the system's vocabulary but not to include a "world" in the AI sense).

Although DLT is claimed to be interlingual, it appears to be "double transfer" because it includes a full translation module between the source language and the interlingua and a full translation module from the interlingua into the target language, complete with parsers, transfer modules, and generators. WITKAM summarizes the differences between this approach and the traditional types of transfer.

The most unusual feature of this project is the decision to use Esperanto as the interlingua. Witkam proposes three criteria for the choice of an interlingua—unambiguity, compactness, and inspectability—and claims that BCE (binary-coded Esperanto) is an optimal choice.

We find this choice questionable for several reasons. First, the advantages of Esperanto over framelike or annotated tree data structures are not obvious. The latter also exhibit unambiguity, compactness, and inspectability and, moreover, are superior to the string representation, such as an Esperanto sentence, from the computational point of view.

Second, the real problem in the choice of a representation language is its expressive power and the extent to which it facilitates inference making, which in MT is the process of augmenting the incoming information with information implied in the input text. In other words, a representation language must facilitate "deep understanding." The only theoretical alternative to this premise is to prove that deep understanding (or at least a fraction of that task) is not necessary for high-quality translations. No such nontrivial result has been shown to date.

The whole process of translation in DLT is, thus, twofold: it involves a (transfer-based) translation from a source language into Esperanto and a (transfer-based) translation from Esperanto into the target language. The efficiency of this approach must be compared with the other two valid alternatives: 1) the conventional transfer approach, and 2) the use of a different kind of interlingua.

If a translation system is aimed at a single pair of languages, the Esperanto-based approach is less efficient than the transfer method because an extra analysis/synthesis step is involved. If the system is multilingual, the efficiency of any interlingua-based approach will improve with the number of languages added. On the other hand, if an interlingua is not constrained formally by the requirements of a natural language such as Esperanto (e.g., linearity, ambiguity, hidden associations), which requires separate parsing and generating stages to extract latent semantic and pragmatic features, it will be more difficult to adapt as efficiently as an interlingua that is based on a system of internal meaning (knowledge) representation of the AI variety.

The linguistic part of the DLT project has been done largely in the feasibility study; major lexical and syntactic structure classes of Esperanto and their coding have been discussed there. An augmented transition network scheme is chosen for the formal specification of the interlingua grammar. Selectional restrictions and case markers are added to the grammar, making it syntactic/semantic.

Although the report proposes DLT's IL design as an "excellent platform for AI enhancements," no specific AI methodology (inferencing capabilities, expert system technology or other knowledge-related matters) is mentioned as an *essential* part of the initial design.

PHRAN-PHRED and MOPTRANS

PHRAN-PHRED and MOPTRANS are two examples of ways to explore the translating capabilities of knowledge-based natural-language understanding systems. The general philosophy of both efforts coincides with the position of CARBONELL ET AL. (see above).

The orientation of these efforts is theoretical, and they are limited-scope experiments in using variants of conceptual-dependency (SCHANK, 1975) representations in the role of the interlingua for MT. No attention has been paid to the problems of a production MT system, so the questions of efficiency, quality of translation, supporting systems, and so forth are not applicable. The reports also contain no discussion of the potential extendability of these systems to wider domains and of implementational difficulties in maintaining and accessing the very large databases that will be necessary for such extensions.

PHRAN (phrasal analyzer) is a parser from natural language into a version of conceptual dependency written by Arens (ARENS; WILENSKY & ARENS). PHRED (phrasal English diction) is a generator written by Upstill, a "language production mechanism" from conceptual-dependency representations into natural language. In the Berkeley AI group led by Wilensky, "PHRAN and PHRED serve as the front and back end to various natural language processing systems. In general, PHRAN and PHRED perform that part of language processing that requires detailed knowledge of the specific language involved" (WILENSKY & MORGAN, p8). An important corollary to this claim is that the understander of natural language need not be "integrated" with the knowledge about the world. Semantics is clearly separated from pragmatics, and it is claimed that linguistic knowledge is sufficient for understanding separate utterances in natural language. It is in this point that this project strategically differs from the MOPTRANS proposal of LYTINEN & SCHANK.

Both PHRAN and PHRED were claimed to be extendable to languages other than English (specifically, Spanish and Chinese). "No changes were made to the actual PHRAN code to tailor it to either of the languages. Moreover, the entire time spent on doing this encoding was about six half-time graduate student months by a student previously unfamiliar with PHRAN" (WILENSKY & MORGAN, p13).

It would be surprising if no attempt at MT were undertaken using the PHRAN-PHRED combination. Indeed, a small system of English-Spanish translation was constructed by MORGAN. Although MT terminology is not used, the approach is interlingual since there is no communication between the processes used by PHRAN and those used by PHRED in a session. The former's output can be entered into the latter since it is encoded in the same way.

At the risk of oversimplification, both PHRAN and PHRED are organized as elaborate pattern matchers working in the space of situation-action rules ("pattern-concept pairs"). For PHRAN the condition part of a rule is a language phrase pattern; the action part is "a conceptual template that represents the meaning of the associated phrase" (WILENSKY & MORGAN, p7). In PHRED the two parts of the rule are interchanged.

An implementation of the MOPTRANS MT system is being developed at Yale University (LYTINEN & SCHANK). In this system another conceptual dependency-based representation scheme is used as the interlingua. The approach is that of memory organization packets (MOPs). "The general idea behind MOPs is to store knowledge which is common to many different situations in only one processing structure, and then to make this processing structure available in all the different situations in which it applies" (LYTINEN & SCHANK, p23). For a comprehensive discussion of the concept of MOPs and related topics see SCHANK (1982a; 1982b). MOPTRANS reads stories about terrorism. The authors claim that "currently, it can read several stories in Spanish and French, and translate these stories into English. We are also working on parsing and generating stories in German, Japanese and Hebrew" (LYTINEN & SCHANK, p28). The report presents an annotated computer listing of a sample session with MOPTRANS.

TRANSLATOR

The present authors have begun a model of MT, called TRANSLATOR, which is philosophically similar to the AI systems discussed above *with one important difference:* we intend to represent the expert knowledge of a human translator and use it in the translation process, thus making the system an experiment in modeling human translation (TUCKER & NIRENBURG). We believe that the time for such an effort has arrived, and resulting translations can be measurably different in quality from those achieved by conventional direct and transfer approaches. We also believe that it would be wasteful to build a knowledge-based MT system that would incorporate linguistic and world knowledge but not the expert knowledge of human translators. One of the extremely rare passages in MT literature devoted to similar problems is the last section in the article by JOHNSON.

In its most general form, our model has the major elements shown in Figure 6. The fundamental distinction between the model and "standard" interlingual or transfer models of MT is the *central* role of the Inspector. The other elements in this model have been developed quite well elsewhere and are not of central interest to our research at present.

According to our proposal, the Inspector is the embodiment of the expert translator. This module examines the alternative (syntactic) parses in IL for

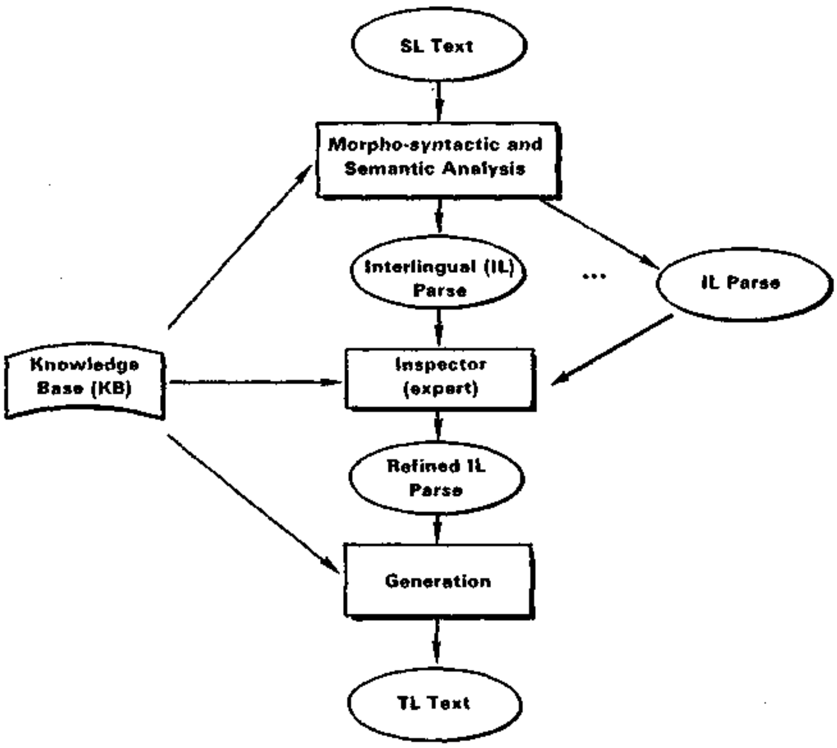

Figure 6. The TRANSLATOR model of machine translation, with interlingua, knowledge base, and expert behavior.

the source text and determines which is the most plausible in the context of its understanding of the translation domain, as represented in the knowledge base (KB). Initially, the Inspector is viewed as the post-editor in conventional MT systems. The challenge here is to simulate the behavior of the post-editor and thus arrive at a high-quality translation *before* the target text is synthesized rather than after. Thus, the Inspector also can assess the completeness and correctness of the final translation as well as its intermediate representation. For example, it will assess how well the pragmatic ambiguity of a text has been resolved, the quality of anaphora resolution, and so forth.

The operation of the Inspector requires the use of a KB. Our first approximation of the form of this KB is shown in Figure 7. It is important that a common representation scheme for these three areas of knowledge be found since the exact boundaries among them are not firm. For example, in the analysis of a specific text, certain "facts" may be discovered that will initially be placed in short-term memory. However, their relevance may extend to future texts in the subject area, in which case these facts should be allowed to "migrate" from short-term memory to long-term memory (where the rest of the world knowledge about that subject area resides). Similarly, the graceful migration of linguistic knowledge to long-term memory and vice versa should also not be prevented.

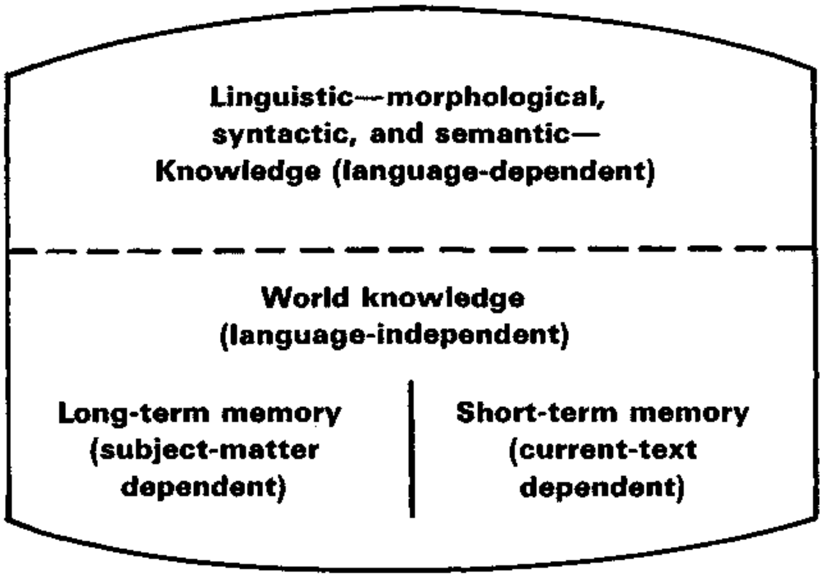

Figure 7. Knowledge-base components of TRANSLATOR

The context in which this design falls is a three-year research project aimed at multilingual MT among four languages: English, Japanese, Russian, and Spanish. Its design philosophy should contribute positively to the art and science of MT as well as to a further convergence between the fields of MT and AI. This project is currently supported by the National Science Foundation.

CONCLUSIONS

R&D in machine translation has found a new vitality and optimism in the past several years. Current systems are productive, but they all require post-editing. New and ambitious projects promise multilingual capabilities by the early 1990s, yet three fundamental problems remain.

First, all the current systems known to these authors are inextricably bound to either the direct or the transfer approach. Some preliminary research has examined the question of interlingua, but no development projects have embraced it in a thorough and convincing way. The design of an effective interlingua for MT continues to be elusive. In our opinion, however, this is an essential element in achieving MT quality that is truly distinguished from that of current production systems.

Second, although AI experiments have demonstrated the potential for machines to "understand" natural language text by incorporating large knowledge bases and inferencing algorithms, their effective adaptation to the specific "world" of MT has not been achieved. Similarly, AI advances in expert systems have not been well adapted to simulate the particular expert behavior of the human translator.

Third, although human translators have an intuitive sense of what makes a good translation and although various efforts have been made to evaluate the quality of MT, no quantitative standard has been defined or demonstrated.

What, then, is the future of MT? Hardware and software tools can now support a high-quality multilingual MT development effort, with a negligible level of post-editing. Such a system cannot be realized until AI researchers, linguists, and computer scientists effectively merge their creativity to achieve this common goal. After all, MT is a common and attractive application area for all three groups. The concept of EUROTRA seems to be a catalyst for this kind of cooperation to succeed. The Japanese Fifth Generation project may also be such a catalyst. We hope that such an effort will soon emerge in the Americas.

BIBLIOGRAPHY

ARENS, YIGAL. 1981. Using Language and Context in the Analysis of Text. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI): Volume 1; 1981 August 24-28; Vancouver, BC. 1981. 52-57.

BAR HILLEL, YEHOSHUA. 1960. The Present Status of Automatic Translation of Language. Advances in Computers. 1960; 1: 91-163. ISSN: 0065-2458.

BATES, MADELEINE. 1978. Theory and Practice of Augmented Transition Network Grammars. In: Bolc, L., ed. Natural Language Communication with Computers. Berlin, West Germany: Springer Verlag; 1978. 191-250. ISBN:3-540-08911-X.

BAUERLE, R., ed. 1979. Semantics from Differing Points of View. Berlin, West Germany: Springer Verlag; 1979. (See particularly chapter by C. Hauenschild, et al.; 324-352).

BEKTAEV, K. B.; BELYAYEVA, L. N.; KOROLIOVA, T. N.; MARCHUK, IU. N.; PIOTROVSKII, R. G.; SADEHIKOVA, P. V.; CHAIKOVSKAIA, I. I.; SHINGAREVA, E. A. 1981. Experiences in Machine Translation of English and Japanese Scientific Texts. Nauchno-Tekhnicheskaia Informatsiia. Seriia 2 (USSR). 1981; (5).

BENNETT, WINFIELD S. 1982. The Linguistic Component of *METAL:* Working Paper LRC-82-2. Austin, TX: University of Texas Linguistic Research Center (LRC); 1982. 24p. (LRC 82-2). Available from: the Center.

BIRNBAUM, LAWRENCE; SELFRIDGE, MALLORY. 1981. Conceptual Analysis of Natural Language. In: Schank, Roger C.; and Riesbeck, Christopher K., eds. Inside Computer Understanding. Hillsdale, NJ: Lawrence Erlbaum Associates; 1981. 318-372. ISBN: 0-89859-071-X.

BOBROW, DANIEL G.; WINOGRAD, TERRY. 1977. An Overview of KRL: A Knowledge Representation Language. Cognitive Science. 1977; 1(1): 3-45.

BOBROW, ROBERT J. 1979. The *RUS* System. 1979. (BBN Report no. 3878). Available from: Bolt, Beranek and Newman (BBN), 50 Moulton Street, Cambridge, MA 02238.

BOBROW, ROBERT J.; WEBBER, BONNIE L. 1980. Knowledge Representation for Syntactic/Semantic Processing. In: Proceedings of the 1st Annual National Conference on Artificial Intelligence; 1980. Los Altos, CA 94022: William Kaufmann, Inc., 95 First Street; 1980. 316-323.

BOITET, CHRISTIAN. 1978. Where Does *GETA* Stand at the Beginning of 1977? In: Overcoming the Language Barrier: Proceedings of the 3rd European Conference on Information Systems and Networks: Volume 1. Munich, West Germany: Dokumentation Verlag; 1978. 88-120.

BOLEY, HAROLD. 1977. Directed Recursive Labelnode Hypergraphs: A New Representation Language. Artificial Intelligence. 1977; 9: 49-85. ISSN: 0004-3702.

BOSTAD, DALE. 1982. Quality Control Procedures in Modification of the Air Force Russian-English MT System. See reference: Lawson, V., ed. 129-134.

BRACHMAN, RONALD. 1978. A Structural Paradigm for Representing Knowledge. Cambridge, MA: Bolt, Beranek and Newman (BBN); 1978. (BBN Report 3605).

BRACHMAN, RONALD. 1979. On the Epistemological Status of Semantic Networks. In: Findler, N. V., ed. Associative Networks. New York, NY: Academic Press; 1979. 3-50.

BRATLEY, PAUL; LUSIGNAN, SERGE. 1976. Information Processing in Dictionary Making: Some Technical Guidelines. Computers and the Humanities. 1976; 10: 133-143. ISSN: 0010-4817.

BRISCOE, E. J. 1983. Determinism and Its Implementation in *PARSIFAL.* See reference: Sparck Jones, Karen;Wilks, Yorick, eds. 61-68.

BRUCE, BERTRAM. 1975. Case Systems for Natural Language. Artificial Intelligence. 1975:6:327-360. ISSN: 0004-3702.

BURTON, RICHARD R. 1976. Semantic Grammar: An Engineering Technique for Constructing Natural Language Understanding Systems. Cambridge, MA: Bolt, Beranek and Newman, Inc. (BBN); 1976. (BBN Report 3453).

CARBONELL, JAIME G.; CULLINGFORD, RICHARD E.; GERSHMAN, ANATOLE V. 1981. Steps Towards Knowledge-Based Machine Translation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1981;PAMI-3 (4): 376-392.

CERCONE, NICHOLAS. 1978. Morphological Analysis and Lexicon Design for Natural Language Processing. Computers and the Humanities. 1978; 11:235-258. ISSN: 0010-4817.

CHARNIAK, EUGENE. 1977. A Framed PAINTING: The Representation of a Common Sense Knowledge Fragment. Cognitive Science. 1977: 1(4): 355-394.

CHARNIAK, EUGENE. 1981. Six Topics in Search of a Parser. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI): Volume 2; 1981 August 24-28; Vancover, BC. 1981. 1079-1087.

CHARNIAK, EUGENE. 1983. Parsing, How To. See reference: Sparck Jones, Karen; Wilks, Yorick, eds. 156-163.

CHARNIAK, EUGENE; RIESBECK, CHRISTOPHER K.; MCDERMOTT, DREW V. 1980. Artificial Intelligence Programming. Hillsdale, NJ: Lawrence Erlbaum; 1980. ISBN: 0-89859.-004-3.

COLMERAUER, ALAIN. 1971. Les SYSTÈEMES-Q; Un Formalisme pour Analyser et Synthetiser des Phrases sur Ordinateur [SYSTEM-Q. A Precise Method for the Analysis and Synthesis of Phrases by Computer]. Université de Montréal, Groupe TAUM; Montréal, Quebec, Canada; 1971. Available from: University of Montreal Dept. of Linguistics.

CORDIER, MARIE O.; MOGHRABI, CHADIA. 1982. An Experiment Towards More Efficient Automatic Translation. In: Proceedings of European Conference on Artificial Intelligence; 1982 July; Orsay, France.

228-231. Available from: the authors, University of Paris, 91405 Orsay, France.

FAHLMAN, SCOTT, 1979. NETL: A System for Representing and Using Real-World Knowledge. Cambridge, MA: MIT Press; 1979. 278p. ISBN: 0-262-0609-8.

FILLMORE, CHARLES J. 1968. The Case for Case. In: Bach, E.; Harms, R., eds. Universals in Linguistic Theory. New York: NY: Holt, Rinehart and Winston; 1968. 1-88.

GOETSCHALCKX, J.; ROLLING, L., eds. 1982. Lexicography in the Electronic Age. Amsterdam, The Netherlands: North Holland; 1982. 276p. ISBN: 0-444-86409-0.

GOLDMAN, NEIL M. 1975. Conceptual Generation. In: Schank, Roger C. Conceptual Information Processing. Amsterdam, The Netherlands: North Holland; 1975. 289-371. ISBN: 0-7204-2507-7.

HENDRIX, GARY G.; SACERDOTI, EARL A.; SAGALOWICZ, DANIEL; SLOCUM, JONATHAN. 1978. Developing a Natural Language Interface to Complex Data. ACM Transactions on Database Systems. 1978 June; 3(2): 105-147. ISSN: 0362-5915.

HUNDT, MICHAEL G. 1982. Working with the Weidner Machine-Aided Translation System. See reference: Lawson, V., ed. 45-52.

HUTCHINS, W. JOHN. 1978. Machine Translation and Machine-Aided Translation. Journal of Documentation. 1978; 34. 119-159.

HUTCHINS, W. JOHN. 1982. The Evolution of Machine Translation Systems. See reference: Lawson, V., ed. 21-38.

ISABELLE, PIERRE. 1984. Machine Translation at the TAUM Group. In: King, Margaret, ed. [Papers from the] International Tutorial on Machine Translation; 1984 April 2-6; Lugano, Switzerland. 59p. Publication forthcoming; for information contact M. King at ISSCO, 54 Route des Acacias, Geneva, Switzerland.

JACKENDOFF, RAY S. 1972. Semantic Interpretation in Generative Grammar. Cambridge, MA: MIT Press; 1972.

JOHNSON, RODERICK L. 1983. Parsing—An MT Perspective. See reference: Sparck Jones, Karen; Wilks, Yorick, eds. 32-38.

JORDAN, S. R.; BROWN, A. F. R.; HUTTON, F. C. 1977. Computerized Russian Translation at ORNL. Journal of American Society for Information Science. 1977; 28 (1).

KAPLAN, RONALD M. 1973. A General Syntactic Processor. In: Rustin, R., ed. Natural Language Processing. New York, NY: Algorithmics Press; 1973. 193-241.

KING, MARGARET. 1980. *EUROTRA:* A European System for Machine Translation. University of Geneva Technical Report. 8p.

KING, MARGARET. 1981. Design Characteristics of a Machine Translation System. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI): Volume 1; 1981 August 24-28; Vancouver, BC. 1981. 42-46.

KING, MARGARET. 1982. EUROTRA: An Attempt to Achieve Multilingual MT. See reference: Lawson, V., ed. 139-148.

KING, MARGARET, ed. 1984. [Papers from the] International Tutorial on Machine Translation; 1984 April 2-6; Lugano, Switzerland. Publication forthcoming; for information contact M. King, at ISSCO, 54 Route des Acacias, Geneva Switzerland.

KITTREDGE, RICHARD; LEHRBERGER, JOHN. 1982. Sublanguage: Studies of Language in Restricted Semantic Domains. Berlin, West Germany: De Gruyter; 1982. 240p. ISBN: 3-11-008244-6.

KNOWLES, FRANCIS E. 1982. The Pivotal Role of the Various Diction-pties in an MT System. See reference: Lawson, Veronica, ed. 149-162.

LAWSON, VERONICA, ed. 1982. Practical Experience of Machine Translation: Proceedings of a Conference; 1981 November 5-6; London, England. Amsterdam, The Netherlands: North Holland; 1982. 199p. ISBN: 0-444-86381-8.

LEHMANN, WINIFRED P.; STACHOWITZ, R. 1972. Development of a German-English Machine Translation System. Austin, TX: University of Texas at Austin Linguistics Research Center; 1972. Available from: the Center.

LEHNERT, WENDY G.; RINGLE, MARTIN H., eds. 1982. Strategies for Natural Language Processing. Hillsdale, NJ: Lawrence Erlbaum; 1982. 533p. ISBN: 0-89859-191-0.

LOH, S. C.; KONG, L. 1979. An Interactive On-Line Machine Translation System (Chinese into English). In: Snell, B. M., ed. Translating and the Computer. Amsterdam, The Netherlands: North Holland; 1979.

LOZINSKII, ELIEZER L.; NIRENBURG, SERGEI. 1982. Parallel Processing of Natural Language and the Locality Phenomenon. In: Trappl, Robert L., ed. Proceedings of the 6th European Meeting on Cybernetics and Systems Research; 1982 April; Vienna, Austria. Amsterdam, The Netherlands: North Holland; 1982. 885-890.

LOZINSKII, ELIEZER L.; NIRENBURG, SERGEI. 1982. The Locality Property of Natural Language and Parsing English. In: Proceedings of the European Conference on Artificial Intelligence (ECAI); 1982; Orsay, France. 216-221. Available from: the authors, Colgate University, Hamilton, NY 13346.

LYTINEN, STEVEN; SCHANK, ROGER C. 1982. Representation and Translation. (Yale AI Project Research Report 234). New Haven, CT: Yale University Computer Science Dept.; 1982. 38p. Available.

MAAS, HEINZ-DIETER. 1984. The MT System *(SUSY)*. In: King, Margaret, ed. [Papers from the] International Tutorial on Machine Translation; 1984 April 2-6; Lugano, Switzerland. 45p. Publication forthcoming; for information contact M. King at ISSCO, 54 Route du Acacias, Geneva, Switzerland.

MARCUS, MITCHELL P. 1980. A Theory of Syntactic Recognition for Natural Language. Cambridge, MA: MIT Press; 1980.

MASTERMAN, MARGARET. 1982. The Limits of Innovation in Machine Translation. See reference: Lawson, V., ed. 163-186.

MCCORDUCK, PAMELA. 1983. An Introduction to the Fifth Generation. Communications of the ACM. 1983; 26(9): 629-631. ISSN: 0001-0782.

MONTAGUE, RICHARD. 1974. Formal Philosophy. New Haven, CT: Yale University Press; 1974.

MORIN, G. 1978. SISIF: Système d'Identification, de Substitution et d'Insertion de Formes [SISIF: A System for Identification, Substitution and Insertion of Forms]. Montreal, Quebec, Canada: Université de Montréal Groupe TAUM; 1978. Available from: University of Montreal Dept. of Linguistics.

MORGAN, MICHAEL. 1982. Spanish *PHRAN* and *PHRED.* 1982. (Unpublished Computer Listing). Information available from: University of California at Berkeley Artificial Intelligence Group.

NAGAO, MAKOTO. 1982. Machine Translation Project of the Japanese Government. Kyoto, Japan: Kyoto University; 1982. 22p.

NAGAO, MAKOTO. 1983. A Survey of Japanese Language Processing, 1980-1982. In: Kitagawa, T., ed. Computer Science and Technologies. Amsterdam, The Netherlands: North Holland; 1983. 64-70.

NAGAO, MAKOTO; TSUJII, J.; MITAMURA, K.; HIRAKAWA, H., KUME, M. 1980. A Machine Translation System from Japanese into English: Another Perspective of MT Systems. Kyoto, Japan: Kyoto University; 1980.

NIRENBURG, SERGEI; ATTIYA, CHAGIT. 1984. Data Model for Artificial Intelligence Applications. In: Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) COMPDEC84 Conference; 1984 April 24-27; Los Angeles, CA.

NIRENBURG, SERGEI; BEN ASHER, YOSEF. 1984. *HUHU:* The Hebrew University Hebrew Understander. 1984. 50p. (Accepted for publication in the Journal of Computer Languages, 1984 January).

NIRENBURG, SERGEI; BEN ASHER, YOSEF; LEVY, JAKOB; MARCOVITZ, SHAUL; SHANI, OMER. 1982. *HUHU: The* Hebrew University Hebrew Understander. SIGART Newsletter. 1982; 79: 61-62.

NIRENBURG, SERGEI; TUCKER, ALLEN. 1984. Knowledge-Based Multilingual Machine Translation. 1984. (Research Report 3-84) Available from: the authors, Colgate University, Hamilton, NY 13346.

NISHIDA, TOYOAKI. 1983. Studies on the Application of Formal Semantics to English-Japanese Machine Translation. (Doctoral dissertation). Kyoto, Japan: Kyoto University; 1983. Available from: Kyoto University Faculty of Engineering, Dept. of Computer Science.

NISHIDA, TOYOAKI; DOSHITA, S. 1982. An English-Japanese Machine Translation System Based on Formal Semantics of Natural Language. In: Horecky, J., ed. Proceedings of the 9th International Conference on Computational Linguistics: COLING-82; 1982; Prague, Czechoslovakia. Amsterdam; The Netherlands: North-Holland; 1982. 276-282.

NISHIDA, TOYOAKI; DOSHITA, S. 1983. An Application of Montague Grammar to English-Japanese Translation. In: Proceedings of the Association of Computer Linguistics (ACL) Conference on Applied Natural Language Processing; 1983 February 1-3; Santa Monica, CA. 156-165. Available from: Don Walker, Stanford Research Institute, Menlo Park, CA.

PETERSON, JAMES. 1980. Computer Programs for Spelling Correction: An Experiment in Program Design. New York, NY; Springer-Verlag, 1980. (Lecture Notes in Computer Science series, Volume 96). ISBN: 0-387-10259-0.

PHILLIPS, BRIAN; HENDLER, JAMES. 1981. Controlling Parsing by Passing Messages. In: Proceedings of the 3rd Annual Meeting of the Cognitive Science Society; 1981; Berkeley, CA. 228-230. Available from: the Society c/o Allen Munroe, University of Southern California Behavioral Technology Lab, 1845 South Elena Ave., Redondo Beach, CA 90277.

PIGOTT, IAN M. 1979. Theoretical Options and Practical Limitations of Using Semantics to Solve Problems of Natural Language Analysis and Machine Translation. In: MacCafferty; Gray, eds. The Analysis of Meaning: [Proceedings of] Informatics 5. London, England: Aslib; 1979. 239-268.

RIEGER, CHARLES. 1975. Conceptual Inference Making. In: Schank, Roger C., ed. Conceptual Information Processing. Amsterdam, The Netherlands: North Holland; 1975. 157-288.

RIEGER, CHARLES; SMALL, STEVEN. 1981. Towards a Theory of Distributed Word Expert Natural Language Parsing. IEEE Transactions on Systems, Man and Cybernatics. 1981; SMC-11(l): 43-51.

RIESBECK, CHRISTOPHER K. 1975. Conceptual Analysis. In: Schank, Roger C. Conceptual Information Processing. Amsterdam, The Netherlands: North Holland; 1975. 83-156.

RIESBECK, CHRISTOPHER K. 1983. Confessions of a Parser Person. See reference: Sparck Jones, Karen; Wilks, Yorick, eds. 178-181.

RITCHIE, GRAEME. 1982. Using AI Methods for Linguistics Ends. In: Proceedings of the Science and Engineering Research Council (SERC) Natural Language Parsing Workshop; 1982 April; Wivenhoe Park, University of Essex. 51-57.

ROLLING, L. N. 1980. Computer-Aided Translation Today and Tomorrow. In: Taylor, P. J., ed. New Trends in Documentation and Information: Proceedings of the International Federation for Documentation (FID) 39th Congress; 1980.

ROLLING, L. N. 1981. The Second Birth of Machine Translation: A Timely Event for Database Suppliers and Operators. Electronic Publishing Review (UK). 1981; 1(3): ISSN: 0260-6658.

SAGER, NAOMI. 1981. Natural Language Information Processing: A Computer Grammar of English and Its Applications. Reading, MA: Addison-Wesley;1981. 399p. ISBN: 0-201-06769-2.

SAWAI, S.; SUGIMOTO, M.; UKAI, N. 1982. Knowledge Representation and Machine Translation. Fujitsu Science and Technology Journal (Japan). 1982; 18(1): 117-133.

SCHANK, ROGER C. 1975. Conceptual Information Processing. Amsterdam, The Netherlands: North Holland; 1975. 374p. ISBN: 0-7204-2507-7.

SCHANK, ROGER C. 1982a. Dynamic Memory. Cambridge University Press; 1982. 234p. ISBN: 0-52124-858-2.

SCHANK, ROGER C. 1982b. Reminding and Memory Organization: An Introduction to MOPs. In: LEHNERT, WENDY G.; RINGLE, MARTIN H., eds. Strategies for Natural Language Processing. Hillsdale, NJ: Lawrence Erlbaum; 1982. 455-494. ISBN: 0-89859-191-0.

SCHANK, ROGER C.; ABELSON, ROBERT. 1977. Scripts, Plans, Goals and Understanding. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977. ISBN: 0-470-99033-3.

SCHANK, ROGER C.; RIESBECK, CHRISTOPHER K. 1981. Inside Computer Understanding. Hillsdale, NJ: Lawrence Erlbaum Associates; 1981. 386p. ISBN: 0-89859-071-X.

SHIAO-SHU, L. 1984. Translation of English into Chinese on a Microcomputer. International Conference on Machine Translation; 1984 February; Cranfield, England. Available from: Allen Tucker, Colgate University, Hamilton, NY 13346.

SLOCUM, JONATHAN. 1980. The LRC Machine Translation System: An Application of Textual and Natural Language Processing Techniques to Translation Needs. 1980 Fall. 18p. (Siemens Corporation Memorandum).

SLOCUM, JONATHAN. 1981. A Practical Comparison of Parsing Strategies for Machine Translation and Other Natural Language Processing Purposes. Austin, TX: University of Texas; 1981. (PhD dissertation). Available from: University Microfilms, Ann Arbor, MI.

SLOCUM, JONATHAN. 1982. A Status Report on the LRC Machine Translation System: Working Paper LRC-82-3. Austin, TX: University of Texas: Linguistic Research Center (LRC); 1982. 14p. Available from: the Center.

SLOCUM, JONATHAN; BENNETT, WINFIELD C. 1982. The LRC Machine Translation System: Working Paper LRC-82-1. Austin, TX: University of Texas, Linguistic Research Center (LRC); 1982. 25p. Available from: the Center.

SMALL, STEVEN. 1981. Viewing Word Expert Parsing as Linguistic Theory. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI); 1981 August 24-28; Vancouver, BC. 1981. 70-76.

SOMERS, HAROLD L. 1983. Investigating the Possibility of a Microprocessor-Based Machine Translation System. In: Proceedings of the Association of Computational Linguistics Conference on Applied Natural Language Processing; 1983 February 1-3; Santa Monica, CA. 1983. 149-155. Available from: Don Walker, Stanford Research Institute, Menlo Park, CA.

SOWA, JOHN F. 1970. A Conceptual Schema for Knowledge-Based Systems. Communications of the ACM. 1970; 23: 193-195.

SPARCK JONES, KAREN; WILKS, YORICK, eds. 1983. Automatic Natural Language Parsing. Chichester, England: Ellis Horwood Ltd.; 1983. 208p. ISBN: 0-85312-621-6.

STEWART, G. 1978. Specialisation et Compilation des ATN: REZO. In: Proceedings of COLING-78; 1978; Bergen, Norway. Available from: the author, University of Montreal, Dept. of Linguistics (Group TAUM), Montreal, Quebec, Canada.

THOMPSON, HENRY. 1983. Natural Language Processing: A Critical Analysis of the Structure of the Field with Some Implications for Parsing. See reference: Sparck Jones, K.; Wilks, Yorick, eds. 22-31.

THOUIN, BENOIT. 1979. The Future of Computer Translation. C I P S Review (Canada). 1979; 3(6).

TOMA, PETER. 1977. *SYSTRAN* as a Multi-Lingual Machine Translation System. In: Commission of European Communities. Overcoming the Language Barrier. Munich, West Germany: Dokumentation Verlag; 1977. 569-581.

TUCKER, ALLEN. 1984. A Perspective on Machine Translation: Theory and Practice. Communications of the ACM. 1984; 27(4): 322-329. ISSN: 0001-0782.

TUCKER, ALLEN; GNUGNOLI, GULIANO; LONG VO, NGUYEN; CHALOUPKA, BEDRICH. 1978. Implementation Considerations for Machine Translation. In: Proceedings of the Association for Computing Machinery (ACM) National Computer Conference; 1978 December 4-6; Washington, DC. 884-890. Available from: ACM Order Dept., 1133 Avenue of the Americas, New York, NY 10036.

TUCKER, ALLEN; NIRENBURG, SERGEI. 1983. Research Proposal in Machine Translation. Hamilton, NY: Dept. of Computer Science, Colgate University; 1983.

TUCKER, ALLEN; VASCONCELLOS, MURIEL; LEON, MARJORIE. 1980. The PAHO Machine Translation System System: Introduction and User's Guide. Washington, DC: Pan American Health Organization; 1980. 135p.

VASCONCELLOS, MURIEL. 1983. Management of Machine Translation Environment. In: Proceedings of an Aslib Conference on Translating and the Computer; 1983. London, England: Aslib; 1983.

VAUQUOIS, B. 1975. La Traduction Automatique a Grenoble. Paris, France: Dunod; 1975.

WEAVER, WARREN. 1955. Translation. Reprinted in: Locke, W. N.; Booth, A. D., eds. Machine Translation of Languages. New York, NY: Wiley; 1955. 15-23.

WILENSKY, ROBERT. 1980. Meta-Planning: Representing and Using Knowledge about Planning in Problem Solving and Natural Language Understanding. Berkeley, CA: University of California at Berkeley, Electronic Research Laboratory; 1980. (Memorandum no. UCB/ERL/ M80/33).

WILENSKY, ROBERT; ARENS, YIGAL. 1980. *PHRAN:* A Knowledge-Based Approach to Natural Language Analysis. Berkeley, CA: University of California at Berkeley, Electronic Research Laboratory; 1980. (Memo UCB/ERL/M80/34).

WILENSKY, ROBERT; MORGAN, MICHAEL. 1981. One Analyzer for Three Languages. Berkeley, CA: University of California at Berkeley, Electronic Research Laboratory; 1981. (Memorandum no. UCB/ERL/ M81/67).

WILKS, YORICK. 1973. An Artificial Intelligence Approach to Machine Translation. In: Schank, Roger C.; Colby, K., eds. Computer Models of Thought and Language. San Francisco, CA: Freeman; 1973. 114-151.

WILKS, YORICK. 1975. An Intelligent Analyzer and Understander of English. Communications of the ACM. 1975; 18. 264-274. ISSN: 0001-0782.

WILKS, YORICK. 1976. A Preferential, Pattern-Seeking Semantics for Natural Language Inference. Artificial Intelligence. 1976; 6: 53-74. ISSN: 0004-3702.

WILKS, YORICK. 1978. Making Preferences More Active. Artificial Intelligence. 1978; 11: 197-223. ISSN: 0004-3702.

WILKS, YORICK. 1979. Machine Translation and Artificial Intelligence. In: Snell, B. M., ed. Translating and the Computer. Amsterdam, The Netherlands: North Holland; 1979.

WILKS, YORICK. 1983. Does Anyone Really Still Believe This Kind of Thing? In: Sparck Jones, Karen; Wilks, Yorick, eds. 182-189.

WINOGRAD, TERRY. 1983. Language as a Cognitive Process: Volume 1, Syntax. Reading MA: Addison-Wesley; 1983. 640p. ISBN: 0-201-08571-2.

WITKAM, A. P. M. 1983. Distributed Language Translation, Feasibility Study of a Multilingual Facility for Videotex Information Networks. Utrecht, The Netherlands: Euro voor Systeemontwikkelung; 1983. [Various paging].

WONG, H. K. T.; MYLOPOULOS, JOHN. 1977. Two Views of Data Semantics: A Survey of Data Models in Artificial Intelligence and Database Management. Information. 1977; 15(3): 344-383.

WOODS, WILLIAM A. 1970. Transition Network Grammars for Natural Language Analysis. Communications of the ACM. 1970; 13: 591-606. ISSN: 0001-0782.

WOODS, WILLIAM A. 1980. Cascaded ATN Grammars. American Journal of Computational Linguistics. 1980; 6(1): 1-12.

XINSONG, J.; GUONING, S.; YU, C. 1983. AI Research in the People's Republic of China: A Review. AI Magazine. 1983 Winter; (4): 43-48.

ZACHARY, WAYNE W. 1979. Translation Databases: Their Content, Use and Structure. In: Raben, J.;Marks,G.,eds. Data Bases in the Humanities and Social Sciences. Amsterdam, The Netherlands: North Holland; 1979.

ZARECHNAK, MICHAEL. 1979. The History of Machine Translation. In: Trends in Linguistics: Studies and Monographs. 1979; 11: 3-87.