Latest Developments in Machine Translation Technology: Beginning a New Era in MT research

John Hutchins
University of East Anglia, Norwich, England

1. MT during the 1980s

Six years ago at the first MT Summit conference, the field of MT was dominated by approaches which had been established in the late 1970s. These were the systems which had built upon experience gained in what may be called the 'quiet' decade of machine translation, the ten years after the publication of the ALPAC report in 1966 had brought to an end MT research in the United States and had profoundly affected its support elsewhere.

Throughout the 1980s, it can be asserted without contradiction, the dominant framework of MT research was the essentially syntax-oriented 'transfer' approach exemplified by such systems as ARIANE at Grenoble University, METAL at Texas, SUSY at Saarbrücken, the Mu system at Kyoto University, and of course the multilingual Eurotra project of the European Communities. In addition, many of the commercial systems which appeared at this time were based on the same principles.

For some time it appeared as if the 'interlingua' approach was not viable. Earlier efforts in the 1970s had been unsuccessful at Grenoble - the CETA system - and at the University of Texas. These were, however, essentially syntax-oriented approaches: while structural transfer was via interlingual ('universal') tree representations, lexical transfer was still via bilingual dictionary substitution. During the 1980s, new approaches to the interlingua model appeared. Some remained essentially linguistic in orientation: the DLT and the Rosetta projects in the Netherlands, but others argued for knowledge-based approaches based on experience in Artificial Intelligence on natural language understanding systems. The most notable centre to emerge was of course at Carnegie-Mellon University.

At the same time, MT was coming out of the laboratory onto the marketplace and into the office. The first milestones came in 1976 with the installation of Meteo for the Canadian broadcasting service and the decision by the Commission of the European Communities to purchase and develop the Systran system - initially for English-French translation, and later for many other Community language pairs. Within five years, ALP Systems, the Weidner company and the Logos Corporation began marketing a wide range of MT systems, and at the Pan American Health Organization the SPANAM and ENGSPAN systems had begun their successful daily production. Shortly, nearly all the Japanese computer companies had begun research on MT systems and many of them put systems onto the market, mainly for English-Japanese translation, during the 1980s. An important consequence was that expectations were more realistic and less utopian than in the pre-ALPAC period. In the 1980s, MT systems were promoted as aids for translation, and their deficiencies were freely acknowledged and accepted. As a result the earlier antagonism of professional translators has diminished.

All this activity stimulated a resurgence in interest in the potential of MT by the general public and by government bodies. One outcome has been the efforts to bring together researchers, users and policy makers in the series of MT Summit conferences and to provide an international forum for information about MT in the establishment of the International Association for Machine Translation. More recently in Europe the Commission of the European Communities has sponsored a series of seminars in member states devoted to the promotion of

the 'language industry' in which MT and translation aids play an important role. A further sign of the revival of government-level interest has been the recent JTEC report for US policy makers about Japanese research activity (Carbonell et al. 1992).

At the end of the late 1980s a number of developments marked changes in the MT research picture, which may well indicate the beginning of a new era in the history of MT activity.

The most distinctive development is the emergence of what is loosely called 'corpus-based' methods and approaches alongside more familiar 'rule-based' methods. In the former are included the statistical methods and example-based approaches, as well as the attention to the use of text corpora as sources of data for linguistic and knowledge databases.

I shall begin by looking at the recent developments in the 'rule-based' framework and then go on to the newer corpus-based methods. This will be, by no means, a complete survey, since it will concentrate on MT research activity and will focus almost exclusively on the changes and differences and largely ignore the continuities: new approaches and new projects will be mentioned at the expense of systems and projects already established before the end of the 1980s. Above all, it will not cover translators aids (except in passing) and it will not deal with the operational and user aspects of MT, since these will be the subject of other presentations at this conference.

Virtually all the developments have taken place since 1989, which I date as the beginning of a new era in MT research. The significance of the date should become clear in the course of this article, but I shall summarise at the end.

2. Rule-based MT

The assumption of rule-based MT is that translation is a process requiring the analysis and representation of the 'meaning' of source language texts and the generation of equivalent target language texts. Representations should be unambiguous lexically and structurally. There have been and remain two major approaches: (a) the 'transfer' approach in which the translation process operates in three stages - analysis into abstract source language representations, transfer into abstract target language representations, and generation or synthesis into target language texts; and (b) the two-stage 'interlingua' model, where analysis is into some language-neutral representation and generation starts from this interlingual representation.

2.1. Transfer-based and interlingua-based MT

Two of the most significant transfer-based projects of the 1980s have now come to an end. These are the Ariane and Eurotra systems which exemplify the typical "second generation" model: batch processing with post-editing and no interactive components, essentially syntax-oriented and stratificational, with analysis and generation passing through series of levels (morphological, syntactic, deep syntactic, semantic), and making little use of pragmatic or discourse information.

However, the projects do 'live on' to a certain extent in the newly established Eurolang project (Roudaud 1992; Pérot 1992). This is based at SITE, a French company which purchased B'VITAL, the Grenoble company founded to develop ARIANE in the French National MT project Calliope. Also involved is the German company Siemens-Nixdorf which is to contribute in some as yet unspecified way with its METAL system. The project is aiming to produce by 1995 a 'user-friendly' MT environment for translators, with facilities for pre- and post-editing, direct access to dictionaries, project management, etc. Initially Eurolang will develop ten language pairs, English into and from French, German, Italian and Spanish, and French into and

from German. The project is to be based on Ariane and METAL and will build upon the expertise of research teams involved in the Eurotra project.

The LMT project which began in the mid-1980s is based at a number of IBM research centres in Germany, Spain, Israel and the United States, with the major prototypes being English-German, German-English, and English-Spanish (McCord 1989; Rimon et al. 1991). Translation is via four steps implemented in Prolog: lexical analysis, producing descriptions of input words and their transfers; syntactic analysis of source texts, producing representations of both surface and deep (logical) relations; transfer, involving both isomorphic structural transfer and restructuring transformations; and morphological generation of target texts. Slot grammar is characterised as combining a lexicalist approach to grammar and logic programming - LMT stands for 'Logic programming MT'.

Although rejecting the 'transfer' model, the interlingua systems of the 1980s were also rule-based. Examples are the DLT and Rosetta projects, both now also at an end. The DLT project adopted a modified form of Esperanto as its interlingua, and the Rosetta system at Philips is well known for the innovative exploration of the isomorphic principle for constructing interlingual representations and the integration of Montague semantics. A number of Japanese projects in the 1980s were also linguistics-based interlingua systems. Among them should be mentioned NEC's PIVOT system, which has been successfully demonstrated for English, Japanese, Korean, French and Spanish (Okumura et al. 1991).

The most important research on the rule-based interlingua approach, however, has been done at Carnegie-Mellon University. From 1985 until 1989, the team worked on a knowledge-based MT system (KBMT89) which has been described in detail in Goodman and Nirenburg (1991). (A major exposition of knowledge-based MT is the monograph by Nirenburg et al. 1992.) The knowledge based approach is founded on the assumption that translation must go beyond linguistic knowledge and must involve 'understanding'. Apart from familiar syntactic analysis and generation components, the KBMT-89 model includes a 'mapping rule interpreter' for converting LFG-type structures into semantically interpreted representations and an interactive 'augmentor' for residual ambiguities. The semantic mapper draws information from a knowledge database of the domain (initially computer manuals). Interlingual representations are intended to convey the 'actual events' of source texts as networks of fully interpreted (expanded) propositions, i.e. events or states with their arguments and causal, temporal, spatial, etc. links to other events or states. An important feature is the representation of anaphoric links, textual relations, speech acts, topic-comment relationships, etc. A number of interlingua models have been developed, notably KANT, and the team has devoted much work on problems of knowledge and lexical acquisition. From this research have come a number of important subsequent projects, notably the Caterpillar and Pangloss projects (see below.)

In recent years a second major American interlingua-based project has been ULTRA at New Mexico State University (Farwell & Wilks 1991). This is multilingual interlingual system for English, Chinese, German, Japanese, Spanish, featuring bidirectional Prolog grammars, relaxation techniques to give 'near miss' translations (for non-standard input), a preference semantics parser, access to large machine-readable dictionaries, and multilingual text editing. The system is claimed to be easily extensible to new languages as source or target, and tolerant of grammars based on a variety of approaches: Semantic Definite Clause Grammar (Spanish), Case grammar, Categorial grammar. Like CMU, the team has paid particular attention to lexical acquisition, primarily from Longman Dictionary of Contemporary English, and have developed a partially automated lexical entry system (Farwell et al. 1992)

A third example from the United States of a knowledge-based interlingua system is the

research at the Microelectronics and Computer Technology Corporation (MCC). Particular emphasis is placed on reversibility and idiomatic generation of target texts (Barnett et al. 1991), and on integrating discourse features into representations (Aone 1991).

2.2. Constraint-based formalisms

A general framework for many linguistics rule-based systems of the present time has emerged clearly within the last five years or so. It embraces all the formalisms which can be categorised as variants or equivalents of 'unification' and 'constraint-based' formalisms (Sadler & Arnold 1992). Earliest chronologically comes Lexical-Functional Grammar (LFG), inspired by work on lexicalist non-transformational linguistics. The Functional Unification Grammar formalism of Kay (1984) has had immense influence on current rule-based MT research. Another strand was logic programming, derived ultimately from the Q-system formalism of the TAUM project, but best known through the Prolog programming language. From this came Definite Clause Grammar, and also such formalisms as the 'slot grammar' found in LMT (Logic programming MT). Other non-transformational grammar research has resulted in the formalisms of Generalized Phrase Structure Grammar, and later Head-driven Phrase Structure Grammar - models which have attracted a number of MT researchers, too numerous to list here.

One of the main stimulants to the exploration of 'constraint-based' formalisms was the Eurotra project. Although it has itself virtually ended, a number of small-scale projects continue to explore formalisms developed by Eurotra researchers: the CAT2 project at Saarbrücken, a transfer-based system employing a formalism developed in Eurotra (Sharp & Streiter 1992), and the MiMo project at the universities of Utrecht and Essex exploring LFG (van Noord et al. 1991).

A major attraction of unification- and constraint-based formalisms is their inherent reversibility - a feature which is assumed to be particularly desirable for bi-directional MT systems. The Rosetta system of the 1980s was the first major investigation; since then, it has been the explicit goal of numerous projects - including LMT (already mentioned), the various Eurotra offspring, the research at ISSCO (below), the CRITTER project, and UNITRAN.

2.3. Principles-based MT

In recent years, there has been MT research in the framework of the principles-and-parameters approach, manifested in syntax as the Government-Binding Theory. The basic premise is there are universal principles which hold across all languages. Distinctions between languages are accounted for by different settings of the parameters, both syntactic and lexical-semantic.

One example is the ITS system under development at Geneva by Erich Wehrli (Wehrli 1992), a PC-based French-English system designed to involve users in interactive analysis.

However, the best known example is UNITRAN. This interlingua-based system has two basic processing components: syntactic processing for the acceptance and production of grammatically correct sentences, and lexical-semantic processing for deriving an underlying conceptual representation and for matching this to appropriate target-language structures (Dorr 1990, Dorr 1992). Both combine language-independent and language-specific knowledge. At the syntactic level these are supplied respectively by the principles and parameters approach of Government and Binding Theory. At the lexical-semantic level, the language-independent information is supplied by a small set of semantico-conceptual primitives, similar in kind to the components of Shank's Conceptual Dependency representations; and the language-specific information is supplied by definitions combining lexical-semantic primitives and language-

specific syntactic constraints, e.g. reflecting the different patterns associated with *like* and German *gefallen*: *I like Mary, Mary gefällt mir*. The advantage claimed for this approach is that separate analysis and generation processors do not have to be written for each language. Going from one language to another is a matter of setting lexical and syntactic switches for each language. Reversibility and bi-directionality are thus readily achieved.

2.4. Lexicalist approaches

A feature of UNITRAN and many 'constraint-based' approaches (including LFG, GPSG and HPSG) is a general movement away from syntax-based representations to more 'lexicalist' approaches.

At its extreme, the essence of the lexicalist approach in MT system design is to reduce transfer rules to simple bilingual lexical equivalences. Such a drastic reduction was first put forward in the CRITTER project (Isabelle et al. 1988). The approach has been explored in the ACQUILEX project devoted primarily to the construction of multilingual lexicons for transfer-based MT (Sanfilippo et al. 1992), and is probably best known in the 'shake-and-bake' method described by Whitelock (1992). The requirement for structural representations - common to both transfer and interlingua approaches - is abandoned in favour of sets of semantic, syntactic constraints on lexical items. Translation involves the identification of target-language lexical items which satisfy the semantic constraints attached to the source-language lexical equivalents. The 'bag' of target lexical items is then 'shaken' to generate output text conforming to the syntax and semantics of the target language.

2.5. General-purpose systems

At the Swiss ISSCO researchers have developed the ELU (Environnement Linguistique d'Unification) as unification shell for natural language processing and in particular for MT experimentation (Estival et al. 1990; Estival 1992). The aim is to provide a theory-neutral framework which facilitates experimentation with alternative linguistic theories and approaches. The framework has been utilized in the development of a essentially transfer-based system for translating avalanche bulletins between French and German (Bouillon & Boesefeldt 1992). The vocabulary and syntax constitute a definable sublanguage suitable for fully automatic translation without requiring post-editing.

The KIELIKONE machine translation workstation (Jäppinen et al. 1991) has also been designed as a general language-independent and theory-neutral MT machine for manipulating general feature tree representations with several in-built inference facilities. Practical implementation at present is a workstation for Finnish-English translation based on dependency-tree representations in a transfer architecture.

The Core Language Engine (CLE), developed at SRI in Cambridge, is a general purpose machine for mapping between natural language sentences and logical form representations, both in analysis and in generation (Alshawi 1992). The aim has been to achieve a substantial coverage of English syntactic and semantic phenomena independent of particular domains. A central feature has been the use of unification as a mechanism for passing information during analysis and generation. The applicability of CLE to MT has been demonstrated on a small English-Swedish bi-directional transfer system (Alshawi & Carter 1990; Alshawi et al. 1992).

Another general-purpose tool which has found MT applications is the PLNLP (Programming Language for Natural Language Processing) project based originally at IBM. Like CLE it is intended as a broad-coverage syntactic and semantic analysis and generation mechanism primarily for English. A number of MT experiments have applied the PLNLP

English Grammar as component in transfer-based systems: English-Japanese (SHALT-2), English-Portuguese (PORTUGA), English-Chinese (C-SHALT), and English-Korean (KSHALT) (Jensen et al. 1993; Shim & Kim 1991).

3. Corpus-based MT

While the new approaches, methods and projects described so far can all be regarded as natural progressions from development having their origins in research of the 1980s or earlier, the emergence of a wide range of what may collectively be called 'corpus-based' approaches and methods represents a new departure in MT research. It is these developments above all which justify the view that MT has entered a new era (see below).

Among the corpus-based approaches we should distinguish between:

a) the direct use of information derived from corpora for the analysis, transfer and generation of translations

b) the indirect use of corpora as sources of information for deriving or compiling lexical, grammatical and knowledge databases, and as sources of statistical information about source and target languages.

In the first group come the statistics-based, example-based and connectionist approaches to MT system design. In the second group come a range of activities concerned with database compilation, lexical and knowledge acquisition and the use of statistical information in otherwise rule-based systems.

3.1. Statistics-based MT

The most dramatic development at the end of the 1980s has been the revival of the statistical approach to MT - not seen since the early 1960s - in the IBM Candide research project (Brown et al. 1988, 1990). A striking feature is the use of stochastic methods as virtually the sole means of analysis and generation. The IBM research is based on a large corpus of the Canadian Hansard, which records parliamentary debates in both English and French. The essence of the method is the alignment of sentences in the two languages and the calculation of the probabilities that any one word in a sentence of one language corresponds to two, one or zero words in the translated sentence in the other language. Alignment is established by a technique widely used in speech recognition. The probabilities are estimated by matching bigrams (two consecutive words) in each English sentence against bigrams in equivalent French sentences.

The success of the approach surprised MT researchers who had been working in the linguistics rule-based tradition. An evaluation test of 73 sentences showed that 48% were either 'exact' (the same as a Hansard translation), 'alternative' (the same meaning but in slightly different words), or 'different' (legitimate translations but not conveying the meaning in Hansard.)

The IBM researchers (Brown et al. 1992) propose to improve performance not only by various statistical techniques (including trigrams as well as bigrams) but also by including data on inflectional morphology and syntactic transformations to deal with discontinuous structures (e.g. *do not want*, *ne veut pas*), to make French structures more like English ones before translation.

Although these modifications may be seen as admissions that purely statistical techniques have their limitations, the approach has undoubtedly stimulated MT researchers to reconsider the fundamentals of their methodologies.

3.2. Example-based MT

The example-based method was first proposed in the mid 1980s (Nagao 1984), but not implemented until the end of the decade. The basic argument is that translation is often a matter of finding or recalling analogous examples, discovering or remembering how a particular source language expression or something similar has been translated before. The development of larger databases with faster access made the computational implementation of the idea feasible. In essence, the method relies on a bilingual database of example phrases derived from a large corpus of texts and their translations.

In its final years, the DLT project in Utrecht examined the approach in some detail. Although not implemented, the proposals by Sadler (1989) have had considerable influence.

The most extensive exploration of the EBMT approach has been in the ATR project known as 'Transfer-Driven MT' (Sumita et al. 1990; Furuse & Iida 1992, 1992a) for spoken language translation. Selection of the most plausible target example translation is based on a calculation of the distance as measured by closeness of semantic attributes in a thesaural hierarchy. Various levels of transfer knowledge are used: string, pattern and grammatical. String transfer involves identification of specific lexical items, pattern transfer the recognition of similarities of constructions allied with measurements of semantic distance, and grammar-level transfer is expressed in terms of grammatical categories. For example, the Japanese N_1 *no* N_2 construction corresponds often to English N_2 *of* N_1. But not always: it is more idiomatic to say *fee for the conference* rather than *fee of the conference*, *conference in Tokyo* rather than *conference of Tokyo*; and literal translations from Japanese such as *holiday of a week* or *reservation of hotel* must be replaced by *week's holiday* and *hotel reservation*. In the TDMT approach, analysis does not aim for the understanding of the structure or the meaning of the input (as in rule-based systems) but the extraction of the information which is necessary for the application of the various levels of transfer knowledge. A lower level of syntactic analysis is also facilitated by the restriction to a rather limited sublanguage (registration for conferences and booking hotel accommodation).

Other examples of the Japanese interest in the example-based method are the research at Hitachi (Kaji et al. 1992) and at Kyoto (Utsuro et al. 1992).

An example of the integration of example-based methods with other approaches is found in the SHALT project at IBM Japan. This is characterised as a knowledge-based multilingual, multi-domain MT project, influenced in a number of respects by the CMU KBMT-89 model. It employs five types of knowledge sources: grammar rules, concept definitions, mapping rules between syntactic and conceptual structures, conceptual paraphrasing rules, and a corpus of example sentences (Takeda et al. 1992). Its grammars are based on LFG formalisms (called pseudo-unification, and originally used in KBMT-89) and the PEG formalism of the PLNLP project. As this brief outline indicates, SHALT incorporates many of the latest developments in MT research.

3.3. Connectionist approaches

One of the most significant developments in the computational modelling of cognition has been research on parallel computation, neural networks or connectionism. A distinctive feature is the computation of the strengths of links between nodes of networks, and the adjustment of the weightings as a result of actual analyses, i.e. the network 'learns' about links and their strengths for later use. Another attractive feature is the possibility of computing alternative analyses in parallel.

Not surprisingly, these developments have attracted MT researchers, although it is only

in the last two years or so that reports have appeared at conferences. The researchers at CMU concerned with speech translation have made a preliminary exploration (Jain et al. 1991.); McLean of UMIST (1992) put forward a connectionist method to compute the distances between input text segments and bilingual text data in an example-based MT model; and Ishikawa & Sugimura of Matsushita (1992) described a neural-network approach in a transfer-based MT prototype. In this, all possible analyses of noun modification structures are computed in parallel. The system uses two types of 'control rules', link control rules defining semantic and syntactic relationships, and level control rules defining precedences of alternatives (e.g. of modifying phrases to verbs). The weights of these links are computed by spreading activation across network nodes, and the preferred analyses are calculated.

Connectionism offers the prospect of systems 'learning' from past successes and failures. Previously, learning has meant that systems suggest changes on the basis of statistics about corrections made by users, e.g. during post-editing. This approach is seen in the commercial Tovna system and in the experimental PECOF 'feedback' mechanism in the Japanese MAPTRAN system (Nishida & Takematsu 1990). A similar mechanism has been incorporated in the PIVOT system from NEC (Miura et al. 1992). However, NEC researchers have also investigated a pattern learning mechanism to supplement rule-based procedures for parsing complex Japanese sentences (Okumura et al. 1992). The module learns to recognise acceptable global sentence patterns from users' corrections. The researchers compared this approach to a fully automatic connectionist 'learning' module, reporting a success rate of 94% on the test corpus.

4. Corpora as sources of information

Research on example-based and statistics-based approaches have emphasised the importance of text corpora in MT research. However, it had already become increasingly urgent for rule-based systems to have access to reliable data. It is reflected in the moves within the general computational linguistics community towards the gathering of large reusable corpora of machine readable texts, dictionaries and lexical databases. Recent major government-supported initiatives are the Linguistic Data Consortium in the United States and the EAGLES (Expert Advisory Group on Language Engineering Standards) in the European Community.

Within MT it is clear that corpus information is essential for a number of purposes. When building sublanguage or domain-specific systems, whether they are prototype systems or about to be marketed, developers must have detailed knowledge of the vocabulary and grammatical features in the kinds of texts the system is intended for (Tsujii et al. 1992).

The building of lexical and knowledge databases is particularly important for interlingua systems with 'conceptual' representations, such as the UNITRAN system and the Knowledge-based systems at Carnegie-Mellon (Gates & Shell 1993), ULTRA (Farwell et al. 1992), the Pangloss project and the LMT project at IBM (Neff & McCord 1990). However, the topic has much wider application, and lexical acquisition and maintenance are becoming frequent topics at MT conferences. Indeed in the last few months there have been two conferences devoted exclusively to the theme: at the AAAI Spring Symposium in Stanford (AAAI 1993) and at an EAMT-organised conference in Heidelberg (Steffens 1993). In these discussions the importance of sharing and reusability of resources is a major theme (e.g. Bläser et al. 1992)

Given the labour-intensive nature of building large dictionaries, it is essential than replication is minimised and cooperative work is promoted. The best known effort here is the Electronic Dictionary Research project in Japan, supported by eight companies involved in MT research. We shall be hearing more about EDR at this conference.

Bilingual text databases are required both for example-based systems or for translator's workstations. A major problem in this area is alignment, i.e. providing the basic links between source and target texts. The best-known research is the area is the work of Church and his team at AT&T (Gale & Church 1991). Research on alignment in the context of a translator's workbench is a major part of the ACQUILEX project (Picchi et al. 1992), and in the work on the Canadian workstation by CWARC (Simard et al. 1992).

5. Direct use of statistical information

Along with the revival of stochastic approaches to MT, there is increasing interest again in the use of statistical information to aid or complement rule-based methods. It was common in pre-ALPAC systems but was not found again until the late 1980s (in the DLT project). An example of the kind of statistical data which can be obtained from corpora is the research at AT&T on the amount of context required for word sense disambiguation in a purely statistics-based approach (Gale et al. 1992)

A number of Japanese projects are employing statistical techniques in conjunction with example-based approaches. A notable example is the use by NEC in its interlingua based system PIVOT (Doi & Muraki 1992) for disambiguation of source language polysemy and for lexical selection in target texts.

The combination of a rule-based approach and considerable use of statistical information is a major feature of the ArchTran system for English-Chinese translation under development in Taiwan (Chen et al. 1991; Su & Chang 1990). Like many researchers, it was found that expansion of the system rule base to a larger scale introduced severe difficulties. The ArchTran 'corpus-based statistics-oriented' model augments a rule-based linguistic database with statistical information providing finer-grained data. For example, the linguistic rules may define permissible cooccurrences of lexical categories, and the corpus-derived statistics the frequencies of conjunctions of particular lexical items.

6. Generation

For most of the history of MT, it has been assumed that the most serious problems were to be found in the analysis and 'understanding' of language: problems of homonyms, structural ambiguity, anaphoric relations. The 'semantic barrier' was regarded as the greatest impediment, i.e. how to identify and represent the specific meanings of all words in texts. The thrust of research on linguistic rules and on knowledge bases reflected this concentration on analysis. Very recently, within the last five years or less, much more attention has been devoted to the question of the generation of fluent target language text.

Recent research within a traditional rule-based conception is illustrated by the work of DiMarco (DiMarco 1990; DiMarco & Hirst 1990) on stylistic improvement of output, in this case French from an English source. The aim is to include a stylistic component in a transfer-based system, to reduce awkward 'literal' output.

The problems of accounting for discourse features have of course continued to attract attention by researchers, e.g. at MCC (Barnett et al. 1991), at Stuttgart (Eberle et al. 1990), at Penang, Malaysia (Mitkov 1992). But the main efforts are in Japan in the area of dialogue translation, and primarily at the ATR laboratories. As already mentioned, this group is exploring various models, example-based MT, LFG, HPSG, etc., within the context of the major aim of this 15 year project: a 1500-word, speaker independent, real-time system with above 75% accuracy, focusing on the domain of conference registration. One approach described by Kudo (1990) is based on the matching of LFG representations against skeletons of 'local context

knowledge'. Another has been looking at illocutionary acts (Kogure et al. 1990). These have been essentially rule-based approaches.

One obvious impetus for the increased interest in generation is the attractiveness of example-based methods and the prospect in the near future of easier availability of the necessary large bilingual text corpora. However, there have been other factors.

6.1. Multilingual generation

One is the need for providing natural language output from searches in databases. Monolingual generation from database representations has been a topic of considerable activity in computational linguistics and artificial intelligence for a number of years, e.g. Bateman & Hovy 1991, Kittredge 1992. Multilingual generation, though equally important, has received less attention. One exception, not surprisingly, is the research of a group long involved in MT. This is the work at Montreal on the FoG component of the RAREAS system for producing marine forecasts in French and English (Bourbeau et al. 1990), and on the LFS system for generating bilingual summaries of statistical data on the labour force (Iordanskaja et al. 1992), applying in the both cases the meaning-text model of Mel'chuk. The important point is that in neither case is there a source 'text' which constrains the style and form of the output. Closely related research on generating meteorological reports is that of Mitkov (1991) In this case the language is Bulgarian, and particular attention is paid to discourse structures.

6.2. Dialogue-based MT

A second impetus comes from the recognition of a different translation need which might be met by MT. It has been a wide-spread assumption that MT systems are for use either by translators, who can post-edit output or interactively improve translations, or by subject expert capable of reading more or less garbled renditions because they know enough of the subject matter. But, there is also a demand for translation from monolinguals who do not know the target language. An example could be the businessman wanting to convey a relatively straightforward message (confirmation of an order, booking of accommodation, etc.) in another language. The MT system can overcome problems of analysis by interactive elicitation with the author of the source 'text' or indeed the system might assist the user to create a text which it can handle. What cannot be done is to ask for assistance with problems of bilingual transfer or of target language generation. The user must be assured that the output is of high quality. Research on what is often called 'dialogue-based MT' is currently being undertaken at UMIST on an English-Japanese system (Jones & Tsujii 1990; Somers 1992), at Brussels (Babel-2) on a system for French-English translation (Jacqmin 1992), at Grenoble (LIDIA) on French into Russian and German (Boitet 1990), and at Kuala Lumpur, Malaysia, on a system for producing official letters in Malay from English (Zaki et al. 1991)

7. Controlled, domain- and user-specific MT

Research on dialogue-based MT brings into focus three closely related significant areas of increased activity in recent years. These are:

a) the control of the language of source texts
b) the restriction of systems to specific domains and/or sublanguages
and (c) the design of systems for specific users.

All three are efforts to overcome in practical operation the known deficiencies of present MT systems - primarily rule-based systems. Since the aims are shared it is often found that combinations of the three options are adopted in the same systems. None of the approaches are

particularly new, of course.

7.1. Controlled input MT systems

Controlled input has been a feature of the Xerox implementation of Systran since the late 1970s; and its success is reflected in its continued use to the present day. The Smart Corporation adopted a similar approach in its customised installations of translation systems for large companies and organisations. In both cases, it was found that control led to improvements in the original source texts. The same experience was reported from the use of a specially devised 'simplified English' with the Weidner MT system at Perkins Engines (Pym 1990)

One of the largest controlled-language projects in future years will be the CATALYST system to be developed by the Carnegie-Mellon team for Caterpillar. A multimillion dollar five-year contract has been signed to produce translation of user and maintenance manuals from controlled English into French, German, Japanese, with minimal post-editing. The system will be based on the knowledge-based interlingua MT system KANT.

7.2. Domain-specific and sublanguage MT

The earliest MT systems were intended ultimately to work with texts from any subject domain and for any environment. This was the aim of most projects in the 1970s and 1980s. Systran, Logos, Ariane, Eurotra, METAL, etc. were intended to be ultimately general-purpose systems. This is true also for most of the Japanese systems of the 1980s, both research and commercial, and the cheaper PC based systems such as the Weidner, Globalink and PC-Translator systems. While it is true that in practice dictionaries and grammatical coverage concentrated on specific subject domains, e.g. computer science, health sciences, etc., this was usually regarded as only the first stage.

In recent years, there has been more emphasis on the specific development of systems for particular domains. It is not an innovation, of course; the idea of the sublanguage goes back many years - to the Meteo system of the mid-1970s. Restriction to a specific domain pervades virtually all small-scale research projects, and many larger ones also. Some have been mentioned already: the ATR project is restricted to dealing with booking hotel accommodation and registering for conferences; the transfer-based CRITTER system covers the sublanguage of weekly reports on the livestock and meat trade in Canada; and the ELU project is devoted to the translation of avalanche warnings.

Examples of larger-scale 'domain-specific' projects include some of the most significant projects recently launched and in progress. Pangloss (Hovy 1993) is a joint project of Information Sciences Institute (Stanford University), Computing Research Laboratory at New Mexico State University and the Center for Machine Translation at Carnegie-Mellon University, for the development of a knowledge-based system for translating newspaper reports on business mergers. The even larger CMU project for Caterpillar (CATALYST) is restricted to the domain of heavy earth-moving machinery. In Japan there is the ATR speech translation project already mentioned, and in Europe there is the recently launched VERBMOBIL project for translating spoken dialogue in international business affairs from German into English, and perhaps eventually Japanese into English (Schulze-Furhoff & Abbou 1992).

7.3. User-specific and custom-built MT systems

A characteristic of many of the systems mentioned in the last few paragraphs is that there were built or are being built for particular customers. It is clearly a growing tendency, and many recent examples of such user-specific or custom-built systems can be given. In most cases,

domains are restricted and the linguistic sophistication is relatively low. In many cases also input texts are controlled. A few examples will illustrate this trend.

In Denmark, the software company Winger has developed since 1988 with the assistance of linguists of the University of Copenhagen a system for translation from English into Danish and Spanish and vice versa. The system integrates MT with word processing, database and communication systems. Recently, the Winger system has been marketed in Europe and the United States.

In Canada, Victor Loewen has developed Le Routier, an English-French and French-English system for translating itineraries for drivers and delivery personnel. The limited vocabulary and grammar and high volume were incentives to developing this Prolog-written system. There is no control of input texts. Le Routier demonstrates how custom-built systems can be developed quickly using well-tried methods for specific purposes: development began in February 1989, and by May 1989 translations were being produced for clients (Loewen 1990).

The Volmac Lingware Services, a software company in Utrecht, the Netherlands, has developed customised domain-specific systems for textile company, an insurance company, for aircraft maintenance manuals, etc. Major features are the editorial control of input texts, standardisation and restriction of vocabulary and sophisticated multilingual text processing facilities. Languages covered so far are: Dutch, English, French, and Spanish (Van der Steen & Dijenborgh 1992).

TRADEX (Aumaitre et al. 1992) translates military telex messages from English into French. It has been developed by Cap Gemini Innovation and Cap Sesa Defense for the French Army. The aim is not high quality translation, but adequate access to basic content. Translations are made either on demand or immediately on receipt of telexes for later consultation. The system uses a unification formalism (Definite Clause Grammar) but goes no further than basic syntactic parsing. There is virtually no semantic disambiguation.

CSK in Japan introduced its own system ARGO in 1988 for Japanese-English translation in the area of finance and economics (Carbonell et al. 1992). Input is specially prepared and there is also post-editing by native English-speaking editors. The system is used by CSK in the monitoring of the Japanese security market, translations are transmitted over the Nikkei Telecom network. ARGO is also offered as a service in Japan, processing over 35,000 pages a year.

HESS (Hangul-English Support System) is a system built for the combined command of Korean and US Forces (Kim et al. 1991). The system is designed for monolingual users wanting translations on demand (database information, map graphics, electronic mail as well as normal documents). It is a transfer model based on LFG formalism.

For obvious reasons, custom-built systems tend not to be innovative theoretically or methodologically. However, they do often exhibit significant technological advances, since the designers are generally more in tune with the requirements of the users and work practices of translators and organisations. Furthermore, the background of many of these developers is in technologically advanced sectors of industry and commerce, with access to the latest computational facilities.

The growth of custom-built systems are a welcome sign that MT methods and techniques are becoming more publicly known outside the narrow circles of MT research. It should be heartening to the researchers that their work is now finding direct and valued application. On the other hand, it is surprising that the early lead given by translators at the Pan American Health Organization in developing SPANAM and ENGSPAN has only recently been followed by others.

8. Commercial systems

Commercial systems tend also to be conservative. In these cases, however, the reason is the need to be robust and reliable, applying well-tested and proven methods. For the present, these will be essentially rule-based and syntax-oriented, predominantly on the transfer-based model. Statistical methods will be absorbed gradually - the ArchTran system, already on the market, may be seen as a forerunner in this respect. How long it will be before the first fully statistics-based commercial system appears is an interesting area of speculation.

The pace of growth and change in the commercial area of MT activity is difficult to monitor. New systems and new language pairs seem to be announced every month. It is worth reminding ourselves that even such well-established systems as METAL, Globalink and PC-Translator were new-comers at the end of the 1980s. (For an overview of availability in the United States see the American Translators Association Report on PC-based MT products, December 1992.)

Since 1989, further companies have joined the field. The market for translation software has been exploited primarily in the United States, usually promoted not as translation systems but as translation aids. In addition to Globalink and PC-Translator there are now: Tovna for English into French, German and Russian; various versions of DP/Translator (from Intergraph, a development of the older Weidner MacroCAT system, with greatly improved interfaces and facilities); Toltran for English-Spanish and Spanish-English; Translate from Finalsoft Corporation; XLT from Socatra (Canada) for English-French translation; an English-Arabic system from AppTek, Virginia; Hypertrans from Italy (developed by Luca D'Agostini); Lexitrans for Russian-English translation, and the various Language Assistant packages from Microtac. Japanese systems since 1989 include: RMT/EJ from Ricoh, DuetQt from Sharp, STAR from Catena (marketed in the US as LogoVista E to J by Language Engineering Corporation), EZ JapaneseWriter, and Meltran from Mitsubishi.

As far as the older mainframe-based systems are concerned, the main developments have been the expansion of installations, the increase in user numbers, and the introduction of new language pairs. For example, use of Systran has increased dramatically in the last few years at the Commission of the European Communities and in NATO (Brussels); the Logos English-French system is being used in the world's largest translation project, at Lexi-tech (Moncton, New Brunswick, Canada); the user base for Siemens' METAL system has increased considerably and research on new language pairs continues; the Japan Centre for Science and Technology has started full production of translations of Japanese abstracts, using its developed version of the Mu system. These are just a few notable examples. I shall not go into further details here as this aspect of MT activity will be treated by other contributors to the conference.

9. The global view

Nearly all the projects and systems described in the preceding paragraphs are based in or originate from North America, Western Europe, and Japan.

However, it is by no means the full picture. The contribution of the multilingual CICC (Center for the International Cooperation for Computerization) project is well known for the collaborative research efforts involving MT groups in China, Malaysia and Thailand. We shall be hearing more about this important project during this conference.

Malaysia is in particular an important centre of MT activity. Mention has already been made of the English-Malay system for official correspondence. A major research project for a number of years has been the JEMAH system inspired by the Grenoble ARIANE approach to MT (Zaharin & Tang 1991).

Information about research activity in China is sparse, but it is reported that many centres have been established for the investigation primarily of English-Chinese translation.

In Korea, there has been a considerable growth of research, often collaborations between Korean institutes or companies and similar bodies in Japan and the United States. Some of these projects have already been mentioned: HESS and KSHALT. Others include MATES (Machine Aided Translation Environment), a transfer-based system for English-Korean translation developed by Systems Engineering Research Institute/Korean Institute of Science and Technology since 1989 (Jang et al. 1991; Lee et al. 1991); an English-Korean knowledge-based system (Kim & Chung 1991); a collaborative project for ECS on a bi-directional
LFG-based English and Korean system (Kim 1991).

The situation in Eastern Europe is not so happy. The recent political developments since 1989 have had major impact on MT research. Central state funding has in many cases evaporated. MT researchers are being encouraged to be entrepreneurs, although without the economic climate to assist them. Many have turned to the conversion of mainframe systems onto personal computers. Examples are the AMPAR system for English-Russian, developed until recently at the Centre for Translation in Moscow, and the PARS system of Blekhman. Some projects have virtually ceased, e.g. the APAC English-Czech project in Prague and much of the research in Leningrad/St Petersburg. However, it is encouraging to learn that the important ETAP-2 project under the direction of Yurii Apresyan is continuing in Moscow.

10. Work stations

Translator's workstations are designed to provide a computer-based environment for professional translators integrating a range of facilities.These may include automated access to dictionaries, terminology databanks and other information sources (on-line remote access, CD-ROM, or stored on a local network, etc.); tools for managing terminological resources; multilingual word processing facilities; optical character recognition devices; electronic receipt and transmission of texts; concordance software; and storage of and access to existing translations for later (partial) reuse or revision, e.g. as sources of example translations or already translated text segments. Researchers in MT have recognized the direct applicability of much of their efforts to these practical needs. The most recent example is the relevance of bilingual text alignment.

Although it is realised that many translators do not want full translation facilities - at most perhaps tools for producing rough 'pre-translations', i.e. replacing terminology consistently and accurately (Bédard 1992) - it is the aim in a number of the projects to include access to MT software.

Many MT research teams have recently been involved in the development of workstations. Among them particular mention should be made of the Canadian Workplace Automation Research Center, which has recently announced its latest prototypes (MTNI #4: 14). Central to the research has been the construction and alignment of bilingual text database, and the development of concordance tools (Simard et al. 1992: 67-81; Isabelle 1992) Another prominent participant has been the Carnegie-Mellon Centre for Machine Translation (Nirenburg 1992), whose workstation will eventually include facilities for example-based translation, post-editing of MT output, and support the interactive 'augmentor' of CMU knowledge-based MT systems. In addition, it is seen as a tool for enhancing knowledge acquisition in the development of MT and other NLP systems (i.e. the ONTOS acquisition tool mentioned earlier).

The integration of MT software features in other workstation projects. The CEC-supported TWB (Translator's Workbench) project may provide remote access to the METAL

translation system (Ahmad & Rogers 1992). In the case of the IBM workstation TranslationManager/2 there are plans to provide access to a translation system based on the LMT research. In Malaysia, a workstation for English-Malay translation is intended to incorporate the JEMAH translation system (Zaharin et al. 1992).

Some major users of MT have built their own workstation environments. The Xerox company, which uses the ALP System as well as Systran, has developed DocuTran, software which combines document publishing and translation, with formatting, editing and revision all at one workstation. A similar software environment for document production and publication has been developed by Krupp Industrietechnik (Freibott 1992)

11. MT and users

MT has now perhaps at last freed itself from the perfectionism of earlier times, or rather from the public assumption that the goal of MT is fully automatic translation. It was this perception which infected attitudes before during and for many years after the ALPAC report. Now MT research knows its limitations, it can explore what is realistic and feasible. The public is more ready to accept the usefulness of the less than perfect. The researcher does not have to attempt to 'mimic' the human translator, he can explore systems which are quite different. For example, simultaneous translation into multiple target languages, and generation from non-textual sources. MT researchers can now free themselves from the assumption that what they are developing are only for translators; they can look at other spheres where translation of some kind has a role, where professional translators are not and would not expect to be involved. In particular, MT can and should turn to the development of tools for the occasional non-professional translator.

Their needs have been virtually neglected. We may identify the following categories: (a) bilingual non-translators, e.g. engineers, doctors, social scientists, administrators, politicians, journalists, etc. who need to read and understand papers and reports in another language; (b) monolingual users, who know only the source language, who may have a clear idea of what they want to say in their own language but are not able to judge whether the translation conveys their message; and (c) monolingual users knowing only the target language, e.g. research scientists wanting to find out what an article says in an unknown language, not necessarily in a stylistically perfect translation, but with assurance that the translation is reliable.

For these non-translators, there is very little available. The cheaper PC-based systems such as PC-Translator and Globalink MTS may help the bilingual user for the occasional translation. They could also be used by monolinguals ignorant of the source language if the input text was already in electronic form and error-free, but these users need to be warned of the dangers of mistranslation. The large mainframe systems (Systran, Logos, and the like) can produce 'rough' versions usable by monolinguals ignorant of the source language as long as they know the subject matter, but this use is an almost accidental by-product of their main purpose.

As we have seen, for monolinguals ignorant of the target language there is some interesting research in progress (sect. above). However, there seem to be no systems being developed specifically for the monolingual user not knowing the source language. The requirements are challenging: no bilingual knowledge can be assumed for interactive analysis or transfer, for correcting OCR input or for understanding garbled output.

12. Evaluation of systems

As MT systems become more widely used, as the attractions of PC-based translation software grow, the need for agreed standards of performance become ever more urgent.

Evaluation of MT systems has emerged as a key issue particularly in the last few years - many would say the concern is long overdue. As far as the MT research community has been concerned the milestones of this activity have been the workshops on evaluation at MT conferences, notably at the 1990 TMI conference. A prime mover has been Margaret King (ISSCO), who chaired the sessions and who subsequently organised an important seminar in April 1991. In November 1992, a sub-committee of the Japan Electronic Industry Development Association published a set of criteria for the economic and technical evaluation of systems by users, and outlined guidance to developers for evaluating their systems (JEIDA 1992).

It is important that the momentum is not lost. For the sake of the credibility of the MT community it is vital that potential users can discover exactly what they may expect. The IAMT can and should play a central role in the establishment and monitoring of MT standards and benchmark tests.

13. Conclusion

When I began I said that we have entered a new era in the history of MT. Evidence for this assertion has been given during the talk. By way of summary, I shall attempt to bring together the evidence.

13.1. The five eras of MT history

The first proposals for using the computer to translate were made by Warren Weaver in March 1947. The 'pioneer' era lasted until January 1954 when a collaboration between IBM and a group at Georgetown University produced the first live demonstration of an MT prototype system. This demonstration stimulated large-scale funding by government agencies in the United States and encouraged the establishment of MT groups throughout the world. A period of intense activity dominated by systems of the "first generation" direct approach, i.e. essentially dictionary-driven with low-level syntactic analysis and use of semantic features, was brought to an end by the publication in 1966 of the ALPAC report. There followed a 'quiet' decade when MT was virtually neglected. A revival came in 1975 and 1976 with the installation of Systran at the Commission of the European Communities and with the public inauguration of the Meteo system in Canada. A decade of increasing activity followed, primarily on indirect systems of the "second generation", i.e. rule-based syntax-oriented stratificational with abstract transfer representations. The dominant paradigm was the 'transfer' approach exemplified by Ariane, Eurotra and Mu. But there was also much work on rule-based interlingua models (DLT, Rosetta and at Carnegie-Mellon University). At same time the first commercial systems appeared, including cheaper PC-based software.

A new era began in 1989 with the emergence of newer approaches: the statistics-based IBM Candide project (Brown et al. 1988), example-based MT (e.g. Sadler 1989), the constraint-based and lexicalist tendencies in rule-based MT, and systems for monolingual use (Boitet 1989, Somers et al. 1990). From about this time on, we have seen also the commercialisation of true "second generation" MT systems (METAL, PIVOT, ATLAS/II, ASTRANSAC, etc.), and the subsequent growth of user-specific custom-built systems based on MT research. Some notable European "second generation" projects have come to an end (Ariane and Eurotra), and MT research in Eastern Europe suffered from the end of Communist power in 1989/1990. On the other hand, a new multilingual European project has begun (Eurolang). Shortly we can expect the first operational implementation of knowledge-based MT (the Pangloss project and CATALYST).

13.2 Towards 'third generation' systems

In this new era we are likely to see the emergence of a "third generation" of MT system design. It will combine the essentials of the rule-based approach of the "second generation" with the 'corpus-based' methods which have come to the fore since 1990. The rule base will be less ambitious than that of the 'indirect' models: syntactic analysis may well be restricted to surface constituency and dependency relations; semantic analysis will probably be limited to identification of sentence and clause roles (agents, patients, etc.) and broad-brush semantic features (human, animal, etc.), for initial disambiguation; and the lexical information will be derived primarily from standard dictionary sources, with 'crude' syntactic categories and semantic 'features'. Transfer will operate on lower-level 'surface' representations (although unlikely to be at as low a level as suggested by the IBM research on statistics-based MT - so reminiscent of earliest direct translation systems) Systems will include combined use of example-based methods (using large bilingual text corpora aligned by simple syntactic category analysis and statistical methods), use of statistical data about lexical collocations and vocabulary frequencies, use of domain-specific knowledge bases (comprising both linguistic and subject knowledge), and use of statistical 'feedback' from the interactions and revisions of actual users. [It is possible that the ATR and ArchTran models of transfer-driven MT may be forerunners of the "third generation" MT design.] At the same time there will be a much greater emphasis on discourse and text stylistic aspects - the shift towards production of idiomatic high quality output is already strong. On the technical side, the "third generation" is likely to be integrated more closely into documentation and publishing systems, as already under way in the development of translator's workstations.

However, the new era will also continue to see commercial exploitation of "second generation" models. The general public will be more aware of MT and its potentials. The MT community does not have much time to establish the standards which potential users need to evaluate the relative merits of systems. If it does not do so, there is a risk that the whole MT endeavour will gain a reputation for exaggerated claims and false promises which is no longer deserved.

14. References

The following citations are only those mentioned directly in the above, which it must again be stressed is not intended to be a complete survey in any sense. The sources for much of this article are to be found in conference proceedings, which are cited as follows:

Coling 88 = Varga, D. (ed) Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics. Budapest: John von Neumann Society for Computing Sciences, 1988.

Coling 90 = Karlgren, H. (ed.) Coling-90: papers presented to the 13th International Conference on Computational Linguistics...1990 Aug, Helsinki, Finland. Helsinki: Yliopistopaino, 1990. 3 vols.

Coling 92 = Boitet, C. (ed.) COLING-92: proceedings of the fifteeth International Conference on Computational Linguistics, 23-28/8/1992, Nantes. Grenoble: GETA, 1992. 4 vols.

MT Summit 3 = Machine Translation Summit III, July 1-4 1991, Washington, D.C. [Pittsburgh: Carnegie-Mellon University], 1991.

TMI-90 = Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June 1990, University of Texas. Austin, TX: Linguistics Research Center, University of Texas, 1990.

TMI-92 = Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, Empiricist vs Rationalist Methods in MT (TMI-92), June 25-27, 1992, Montréal, Canada. Laval (Québec): CWARC, 1992.

FGNLP = Ananiadou, S. (ed.) International workshop on fundamental research for the future generation of natural language processing (FGNLP) proceedings, 30-31 July 1992. Manchester: Centre for Computational

Linguistics, UMIST, 1992.

ISKIT '92 = International Symposium on Natural Language Understanding and AI (NLU & AI), as part of International Symposia on Information Sciences (ISKIT '92), July 12-15, 1992, Kyushu Institute of Technology, Iizuka, Fukuoka, 1992.

NLPRS 91 = Natural Language Processing Pacific Rim Symposium (NLPRS '91); November 25-26, 1991. Singapore: Japan-Singapore AI Center, 1991

USM conference = Proceedings International Conference on Current Issues in Computational Linguistics, 12-14 June 1991, Universiti Sains Malaysia, Penang, Malaysia.

Of particular relevance to the theme was the fourth TMI conference (TMI-92) which was devoted to a 'confrontation' between advocates of traditional rule-based approaches (the 'rationalists') and the newer corpus-based approaches (the 'empiricists').

AAAI (1993): Building lexicons for machine translation: Papers from the 1993 AAAI Spring Symposium, March 23-25, Stanford, California. Menlo Park, Ca.: AAAI Press.

Ahmad, K. and Rogers, M. (1992): Terminology management: a corpus-based approach. In: *Translating and the computer 14* (London: Aslib), 33-44

Alshawi, H. (1992), ed. *The Core Language Engine*. Cambridge, Mass.: MIT Press.

Alshawi, H. et al. (1992): English-Swedish translation dialogue software. In: *Translating and the computer 14* (London: Aslib), 165-174.

Alshawi, H. and Carter, D. (1990): Interactive translation using quasi logical forms. In: McNaught, J., ed. *Proceedings of a workshop on machine translation, 2-3 July 1990*, UMIST, Manchester.

Aone, C. (1991): Classifiers and an interlingua for multilingual machine translation. In: USM conference, 198-210.

Aumaitre, J. M. et al. (1992): TRADEX, un système de traduction de telex. *Meta* 37 (4), 624-634.

Barnett, J. et al. (1991): Capturing language-specific semantic distinctions in interlingua-based MT. In: MT Summit 3, 25-32.

Bateman, J.A. et al. (1991): Towards an architecture for situated text generation. In: USM conference, 289-302.

Bédard, C. (1992): La prétraduction automatique, outil de productivité et d'évolution professionnelle. *Meta* 37(4): 738-760.

Bläser, B. et al. (1992): A reusable lexical database tool for machine translation. In: Coling 92 (2), 510-516

Boitet, C.(1990): Towards personal MT: general design, dialogue structure, potential role of speech. In: Coling 90 (3), 30-35.

Bouillon, P.; Boesefeldt, K. (1992): Problèmes de traduction automatique dans le sous-langage des bulletins d'avalanches. *Meta* 37(4), 635- 646.

Bourbeau, L. et al. (1990): Bilingual generation of weather forecasts in an operations environment. In: Coling 90 (1), 90-92.

Brown, P. et al. (1988): A statistical approach to language translation. In: Coling 88, 71- 76. Also*: Computational Linguistics* 16 (1990), 79-85.

Brown, P.F. et al. (1992): Analysis, statistical transfer, and synthesis in machine translation. In: TMI-92, 63-100.

Carbonell, J. et al. (1992): JTEC Panel report on machine translation in Japan. Baltimore, MD: Japanese Technology Evaluation Center, Loyola College in Maryland; 1992 Jan.

Chen, S.C. et al. (1991): ArchTran: a corpus-based statistics-oriented English-Chinese machine translation system. In: MT Summit 3, 33-40.

DiMarco, C. (1990): *Computational stylistics for natural language translation*. Toronto: Computer Systems Research Institute.

DiMarco, C. & Hirst, G. (1990): Accounting for style in machine translation. In: TMI-90, 65-73.

Doi, S. & Muraki, K. (1992): Translation ambiguity resolution based on text corpora of source and target languages. In: Coling 92 (2), 525-531.

Dorr, B. J. (1990): A cross-linguistic approach to translation. In: TMI-90, 13-32.

Dorr, B. (1992): Parametrization of the interlingua in machine translation. In: Coling 92 (2), 624- 630.

Eberle, K. et al. (1992): Contextual constraints for MT. In: TMI-92, 213-224.

Estival, D. (1992): ELU, un environnement d'expérimentation pour la TA. *Meta* 37(4), 693-708

Estival, D. et al. (1990): A syntax and semantics for feature-structure transfer. In: TMI-90, 131-143.

Farwell, D. et al. (1992): The automatic creation of lexical entries for a multilingual MT system. In: Coling 92 (2), 532-538.

Farwell, D. & Wilks, Y. (1991): ULTRA: a multilingual machine translator. In: MT Summit 3, 19-24.

Freibott, G. P. (1992): Computer aided translation in an integrated document production process. In: *Translating and the computer* 14 (London: Aslib), 45-66.
Furuse, O. & Iida, H. (1992): An example-based method for transfer-driven machine translation. In: TMI-92, 139-150.
Furuse, O. & Iida, H. (1992a): Transfer-driven machine translation. In: FGNLP, 95-111.
Gale, W. A. & Church, K.W. (1991). A program for aligning sentences in bilingual corpora. In: *29th Annual Meeting of the Association for Computational Linguistics: proceedings of the conference, 1991 Jun 18, Berkeley, California*. 177-184.
Gale, W.A. et al. (1992): Using bilingual materials to develop word sense disambiguation methods. In: TMI-92, 101-112.
Gates, D.M. & Shell, P. (1993): Rule-based acquisition and maintenance of lexical and semantic knowledge. In: *Sixth Conference of the European Chapter of the Association for Computational Linguistics, April 1993, Utrecht*. 149-157.
Goodman, K. & Nirenburg, S. (eds.): *The KBMT project: a case study in knowledge-based machine translation*. San Mateo, Ca.: Morgan Kaufmann, 1991.
Hovy, E. (1993): Pangloss: interlingua vivat? [interview with Eduard Hovy] *Language Industry Monitor* no.14, 10-11.
Iordanskaja, L. et al. (1992): Generation of extended bilingual statistical reports. In: Coling 92 (3), 1019-1023.
Isabelle, p. et al. (1988): CRITTER: a translation system for agricultural market reports. In: Coling 88, 261-266.
Isabelle, P. (1992): La bi-textualité: vers une nouvelle génération d'aides à la traduction et la terminologie. *Meta* 37(4), 721-737.
Ishikawa, M. & Sugimura, R. (1992): Natural language analysis using a network model: modification deciding network. In: TMI-92, 55-66.
Jacqmin, L.(1992). La traduction automatique au service de l'utilisateur monolingue. *Meta* 37(4), 610-623.
Jain, A.N. et al. (1991): Connectionist and symbolic processing in speech-to-speech translation: the JANUS system. In: MT Summit 3, 113-117.
Jang, M. et al. (1991): Korean generation in MATES/EK. In: NLPRS 91, 170-177.
Jäppinen, H. et al. (1991): KIELIKONE machine translation workstation. In: MT Summit 3, 107-111.
JEIDA (1992): JEIDA methodology and criteria on machine translation evaluation. Tokyo: Japan Electronic Industry Development Association.
Jensen, K. et al. (1993): Natural language processing: the PLNLP approach.Boston: Kluwer.
Jones, D. & Tsujii, J. (1990: Interactive high-quality machine translation for monolinguals. In: TMI-90, 43-55.
Kaji, H. et al. (1992): Learning translation templates from bilingual text. In: Coling 92 (2), 672-678.
Kay, Martin. Functional unification grammar: a formalism for machine translation. In*: 10th International conference on computational linguistics, Proceedings of Coling84;* 75-78.
Kim, J.J.R. (1991): Bracketing paradoxes in natural language processing. *Ohak Yon'gu: Language Research* 27(1), 99-117
Kim, J.H. & Chung, H.S. (1991): Adjunction structure formalism for English-Korean machine translation system. *Ohak Yon'gu: Language Research*. 27(1), 69-97.
Kim, S. et al. (1991): Machine translation in an information management system. *Ohak Yonku: Language Research* 27(2), 273-286.
Kittredge, R. (1992), ed.: Special issue on text generation. *Machine Translation* 7 (1-2).
Kogure, K. et al. (1990): Illocutionary act based translation of dialogues. In: TMI-90, 47-55.
Kudo, I. (1990): Local cohesive knowledge for a dialogue-machine translation system. In: Coling 90 (3), 391-393.
Lee, H. et al. (1991): The English to Korean transfer in MATES/EK. In: NLPRS '91, 46-52.
Loewen, V. (1990): Le Routier: a machine-translation system using Prolog. In: Canadian Information Processing Society/Canadian Advanced Technology Association. Congress 90. Proceedings, Ottawa, Ont., 57-66.
McCord, M. C. (1989): Design of LMT: a Prolog-based machine translation system. *Computational Linguistics* 15, 33-52.
McLean, I.J. (1992): Example-based machine translation using connectionist matching. In: TMI-92, 35-43.
Mitkov, R. (1991): Generating public weather reports. In: USM conference, 372- 378.
Mitkov, R. (1992). Discourse-based approach in machine translation. In: ISKIT '92, 225-230.
Miura, M. et al. (1992): Learning mechanism in machine translation system "PIVOT". In: Coling 92, (2), 693-699.
Nagao, M. (1984): A framework of a mechanical translation between Japanese and English by analogy principle. In: Elithorn, A. & Banerji,R.(eds.) *Artificial and human intelligence* (Amsterdam: North-Holland, 1984),
Neff, M.S. & McCord, M.C. (1990): Acquiring lexical data from machine-readable dictionary resources for machine translation. In: TMI-90, 85-90.
Nirenburg, S. (1992): Tools for machine-aided translation: the CMU TWS. *Meta* 37(4), 709-720.
Nirenburg, S. et al. (1992): *Machine translation: a knowledge-based approach*. San Mateo, Ca.: Morgan Kaufmann.

Nishida, F. & Takamatsu, S. (1990): Automated procedures for the improvement of a machine translation system by feedback from postediting. *Machine Translation* 5(3), 223-246.
Okumura, A. et al. (1991): Multi-lingual sentence generation from the PIVOT interlingua. In: MT Summit 3, 67-71.
Okumura, A. et al. (1992): A pattern-learning based, hybrid model for the syntactic analysis of structural relationships among Japanese clauses. In: TMI-92, 45-54.
Pérot, J. J. (1992): Eurolang project (European languages). In: *Translating and the computer* 14 (London: Aslib), 103-106.
Picchi, E. et al. (1992): A translator's workstation. In: Coling 92 (3), 972-976.
Pym, P.J. (1990): Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. In: Mayorcas, P.(ed.) *Translating and the computer* 10 (London: Aslib), 80-96.
Rimon, M. et al.(1991): Advances in machine translation research in IBM. In: MT Summit 3, 11-18.
Roudaud, B. (1992): La traduction automatique: l'ordinateur au service des traducteurs. *Meta* 37(4), 828-846
Sadler, L. & Arnold, D. (1992): Unification and machine translation. *Meta* 37(4), 657-680.
Sadler, V. (1989): *Working with analogical semantics: disambiguation techniques in DLT*. Dordrecht: Foris.
Sanfilippo, A. et al. (1992): Translation equivalence and lexicalization in the ACQUILEX LKB. In: TMI-92, 1-11.
Schulze-Furhoff, W. & Abbou, A. (1992): Aspects de l'ingénierie linguistique en Allemagne. *La Tribune des Industries de la Langue* nos 7-8, 18-24.
Sharp, R. & Streiter, O. (1992): Simplifying the complexity of machine translation. *Meta* 37(4), 681-692.
Shim, K. & Kim, Y.T. (1991): The general system design of KSHALT. *Ohak Yonku/Language Research*. 27(2), 189-201.
Simard, M. et al. (1992): Using cognates to align sentences in bilingual corpora. In: TMI-92, 67-81.
Somers, H.L. (1992): Interactive multilingual text generation for a monolingual user. In: TMI-92, 151-161.
Somers, H.L. et al. (1990): Machine translation without a source text. In: Coling 90, (3), 271-276.
Steen, G. van der & Dijenborgh, B.J. (1992): Online correction and translation of industrial texts. In*: Translating and the computer* 14 (London: Aslib), 135-164.
Steffens, P. (1993): Machine translation and the lexicon. *MT News International* no. 5, 2-3
Su, K.Y. & Chang, J.S. (1990): Some key issues in designing MT systems. *Machine Translation* 5(4), 265-300.
Sumita, E. et al. (1990): Translating with examples: a new approach to machine translation. In: TMI-90, 203-212.
Takeda, K. et al. (1992): Shalt2 - a symmetric machine translation system with conceptual transfer. In: Coling 92 (3), 1034-1038.
Tsujii, J.I. et al. (1992). Linguistic knowledge acquisition from corpora. In: FGNLP, 61-81.
Utsuro, T. et al. (1992): Lexical knowledge acquisition from bilingual corpora. In: Coling 92 (2), 581- 587.
Van Noord, G. et al. (1991): An overview of MiMo2. *Machine Translation* 6(3), 201-214.
Wehrli, E. (1992): The IPS system. In: Coling 92 (3), 870-874.
Whitelock, P. (1992): Shake-and-bake translation. In: Coling 92 (2), 784- 791.
Zaharin, Y. et al. (1992): On the use of translator workstations and user- driven MT: towards meeting today's translation needs in Malaysia. In: ISKIT '92, 237-244.
Zaharin, Y. & Tang, E.K. (1991): String-tree correspondence grammars as a base for the automatic generation of analysis programs in machine translation. In: USM conference, 127-136.
Zaki Abu Bakar, A. et al.(1991): Malay official letters translation. In: USM conference, 413-420.