[Presented at conference on ‘New directions in machine translation’, Budapest, 18-19 August 1988.]

Recent developments in Machine Translation
a review of the last five years

W.John Hutchins
(University of East Anglia, Norwich)

I. General overview

Ten years ago MT was still emerging from the decade of neglect which succeeded the ALPAC report. The revival of MT can be attributed to a number of events in the mid 1970's: the decision in 1976 by the Commission of the European Communities (CEC) to purchase the Systran system for development of translation systems for languages of the Community; the first public MT system (METEO) for translating weather reports from English into French; the beginnings of AI-oriented research on natural language processing including MT; and, in particular, the appearance of the first commercial systems ALPS and Weidner, followed shortly by Logos. It was perhaps these systems, however crude in terms of linguistic quality, which more than anything else alerted the translation profession to the possibilities of exploiting the increasing sophistication of computers in the service of translation. By the early 1980's there were again research projects in the United States, Japan and Europe.

Yet the greatest expansion of MT activity has occurred in the last five years; the level of global MT activity has probably reached, if not exceeded, the highest levels during the mid 1960's at the time of the ALPAC report. Interest has grown steadily in Europe and again in the United States, but undoubtedly the greatest surge has occurred in Japan. "A rough guess would indicate that 800-900 people are presently engaged in research and development of MT systems in Japan" (Sigurdson & Greatrex 1987). The largest groups in Europe and the United States (those connected with Systran, the projects at Grenoble (GETA), Saarbrücken (SUSY, ASCOF), Utrecht (DLT) and Eindhoven (Rosetta), and the CEC Eurotra project) probably involve in total about 250-300 persons. Non-Japanese commercial systems may account for another 150-200, and all other projects worldwide may well involve no more than a further 400. In numerical terms the dominance of Japan is clear.

The largest growth area has been in the marketing and sale of commercial MT systems, many for personal computers, and in the provision of MT-based services. The picture in MT research has also changed: in the early 1980's it was still concentrated largely in well-established projects at universities (Grenoble, Saarbrücken, Montréal, Texas, Kyoto, and the Eurotra project) and in connection with systems such as Systran, Logos, ALPS and Weidner; although these centres continue (except for the TAUM project in Montreal), most growth in the past five years has been in MT research supported by commercial companies (generally in the area of computer manufacture and software development), and in small-scale projects by individuals and small groups (often microcomputer-based AI-oriented experimental models).

This paper is an attempt to survey the field in this period and to provide, in effect, an updating of the historical review completed in late 1984 and early 1985 (Hutchins 1986). The aim is to give a general picture of the recent developments of systems and projects already established at that time and to document the emergence of new systems and projects worldwide. Descriptions of individual systems are given in section II. This part outlines the main issues and lines of development at the present time. There is no claim of completeness, but it is hoped that all significant activities have been noticed.

Together with the increased activity there has been a marked growth in publication in recent years. Monographs devoted exclusively to MT topics were rare in the 1970's; since the mid 1980's they have become frequent (e.g. Blatt et al.1985, Bennett et al. 1986, Hutchins 1986, Papegaaij 1986, Luckhardt 1987, Schubert 1987, Goshawke et al. 1987, Lehrberger & Bourbeau 1988). In addition there has been growth in publication of the proceedings of conferences devoted exclusively or partially to MT and MT-related topics (e.g. Coling 84, Picken 1985, Coling 86, Batori & Weber 1986, Picken 1986, King 1987, Nirenburg 1987, Wilss & Schmitz 1987, Picken 1987, Slocum 1988). Other collections of articles have appeared edited by Batori & Weber (1986) and by Slocum (1988); the latter of papers which had previously appeared in Computational Linguistics, the foremost journal in its field. The recent successful foundation of a journal devoted to MT (*Computers and Translation*, since 1986) and the regular appearance of MT items in journals such as *Lebende Sprachen*, *Meta*, *Sprache und Datenverarbeitung*, *Language Monthly* and the recently founded *Language Technology*, all are evidence of vigour and widespread interest. Apart from the monographs mentioned above, recent general surveys of the current scene and future prospects are Slocum (1985/1988), Lewis (1985), Tsujii (1986), and Tucker (1987).

1. Classification of MT system types

The first main distinction is in terms of overall strategy: whether translation from source language (SL) to target language (TL) takes place in a single stage ('direct translation'), in two stages (via an 'interlingua') or in three stages (the 'transfer' approach).

Most of the earliest systems were based on the 'direct translation' approach: systems were designed specifically for a particular language pair, analysis of the source language was limited to problems arising specifically in the translation into a particular target language, so that lexical items and structural features which were considered directly equivalent were not subjected to any syntactic or semantic treatment; where structural transformations (reordering of word sequences) were necessary or where multiple choices of lexical items occurred, the most common approach was to augment the relevant SL dictionary entries with contextual information or with directions for structure changes. SL analysis in the direct approach was thus strictly TL-oriented and dictionary-driven. Direct translation systems suffered from monolithic programs inextricably confusing analysis and generation processes, linguistic data, grammatical rules and programming instructions, and above all from the lack of explicit theoretical foundations.

The interlingua approach largely grew out of dissatisfaction with the perceived inadequacies of the 'direct' approach; but not entirely, for it had long been argued that MT of good quality would only come from translation of the 'meaning' of texts and this implied translation via a (universal) conceptual representation. An additional argument had also been that multilingual MT would be most effective (and most economic) if all languages in the system were translated into and from a single intermediary representation. Various suggestions for interlinguas have been put forward, from the creation of 'logical languages' to the adoption of existing natural languages or artificial languages, and these different approaches continue to the present.

The transfer approach arose primarily in the light of experience with research on interlingua systems. A number of problems had been encountered. Firstly there was the difficulty of establishing interlingual elements, even where only two languages were involved. There was some success with basic syntactic equivalences (by analysis into logical forms), but little with lexical (conceptual) equivalences. Secondly, it was found that in the process of abstraction to language-independent representations too much information was lost about text-oriented structure, e.g. whether a particular expression was the theme (subject) or rheme (comment) of a sentence. As a result, the TL output was often incoherent. In the transfer approach the intermediary representations are not intended to be language-independent; rather SL analysis is into a SL-oriented abstract representation and TL generation is from a TL-oriented representation, and between the interfaces is a SL-TL transfer component. The 'abstractness' the interfaces differs from one system to another, as does the nature and amount of different types of information (syntactic, semantic, pragmatic).

The second main distinction concerns the nature of the relationship between the mechanised translation processes and the operators and users of the systems. Systems can be 'fully automatic' in that no human intervention occurs between input of the text in one language and output of the text in another language, i.e. translation is a 'batch' process. Alternatively, systems can be 'interactive' involving human collaboration during translation processes. Further differences within these two basic approaches are also made, so that now there is a wide variety of system types.

Within the 'fully automatic' systems a distinction can be made between systems which are designed to accept input texts only of a particular subject domain (e.g. nuclear physics) or of a particular text type (e.g. patents) and systems which do not impose such restrictions. At present, probably only Systran (sect.1 below) is capable of tackling, in principle, any text in most subjects - and this is primarily by virtue of the large dictionaries built up over many years. In practice, most MT systems are limited to particular 'sublanguages', the vocabulary (and structures) of a specific subject field. This may be regarded by the designers as an initial constraint, to be removed as the system develops into new subject fields. However, in some cases, the sublanguage limitation has been a deliberate design decision - cf. the Smart systems (sect.3), the Johns Hopkins system (sect.11), SEMSYN (sect.16), the ISSCO project (sect.22), TITRAN (sect.26), etc.

The revision of MT output is now accepted universally as inevitable, given the limitations of computational processing of natural language. It is an accepted feature of almost any large translation service or agency that human translations are revised, and so the 'post-editing' of MT is equally acceptable. However, the revision of MT output is quite different from the revision of human translation; in the latter case, it is usually a matter of stylistic refinement and checking for consistency of terminology. In the case of MT, consistent translation of terminology is easily achieved; what is involved is a good deal of low-level correction of 'simple' grammatical mistakes which no human translator would commit: wrong choices of pronouns, prepositions, definite and indefinite articles, etc. However, much can be done to simplify the editing facilities for post-editors, as demonstrated in the PAHO environment (sect.2 below). For certain purposes (e.g. for information acquisition), unedited or lightly revised versions are acceptable. At the other extreme, where high quality is demanded, the MT output may be regarded as a 'pre-translation' which human translators can use as a rough guide to their own version.

The low quality output of 'fully automatic' systems has encouraged the adaptation of texts to the limitations of MT systems. This can take two forms. In the 'pre-editing' of texts, expressions and constructions are changed to those which it is known the system can deal with - or alternatively, marks are included to indicate particular usages (e.g. indicating for light whether it is an adjective, a noun, or a verb). A more common practice at present is the restriction of input to a constrained language (restricted in vocabulary, in order to avoid homonyms and polysemes, and restricted in the range of syntactic structures). One example is the use of Multinational Customized English by the Xerox Corporation for documents which are to be translated by Systran into a number of languages. Another example is found in the systems developed by Smart (sect.3 below) where documents are composed in an English which is less ambiguous and vague, is easier for foreigners to understand, and is more easily translated in a MT system. It is of course possible for systems to be both designed for a particular sublanguage and to be constrained in vocabulary and syntax.

It was the limitation of MT systems also which encouraged the development of 'interactive' systems. The involvement of human 'assistants' may take place at any stage of translation processes: during analysis, during transfer, during generation, or any combination of these. From developments in the past few years it is evident that we can identify the following types of interaction:

(a) interactive analysis: assistance in the interpretation of input text, principally the disambiguation of polysemes and complex syntactic structures.

(b) interactive rewriting: computer-initiated requests to an author to reformulate input text in a form which the program can deal with (i.e. this is effectively interactive pre-editing) - cf. NTRAN project (sect.20 below)

(c) interactive composition: the author composes text which the computer simultaneously attempts to analyse and translate (i.e. this overlaps with the notion of restricted language input) - cf. CMU projects (sect.10 below)

(d) interactive transfer: assistance with the selection of TL equivalences; this may involve 'disambiguation' if there are for a single SL expression more than one TL option. (The SL form may not be monolingually ambiguous, but only in the context of a given TL.)

(e) interactive generation: assistance in producing fluent output, e.g. selection of appropriate constructions in context (thematisation, topicalisation, etc.). There are no known examples at present; the process would overlap with interactive post-editing.

None of these are mutually exclusive; in fact many systems combine several types. In addition, of course, interactive systems can be subject to the same set of design criteria that apply to 'batch' systems: they can be limited to particular sublanguages, they can be restricted in vocabulary and syntax, their input may be pre-edited (as in many Japanese systems), their output may be post-edited to various levels (for high quality products, for 'information-only' purposes). In overall strategy, interactive systems can be based on the 'direct translation' approach, or on transfer or interlingua approaches, and they can operate in a bilingual or a multilingual environment. Various combinations will be found in the survey below. Nevertheless, the great variety of interactive possibilities underlines the argument that the exploration of all possible machine-aided translation options has scarcely begun.

2. MT and understanding

The major theoretical issue which faces all MT researchers is the place of artificial intelligence in MT systems: briefly, how much 'understanding' of texts is necessary for translation. There are many systems under development which incorporate AI methods involving knowledge databases, semantic networks, inference mechanisms, etc. Some prominent examples are the ASCOF system at Saarbrücken (sect.15), DLT at Utrecht (sect.18), the Translator project (sect. 10), the NTT's LUTE system (sect.36), and the research at ETL in connection with the Japanese ODA project (sect.28). The notion of grafting AI techniques onto more traditional approaches has also been proposed (e.g. expert systems in the GETA project - Boitet & Gerber 1986.) By contrast, some researchers have deliberately eschewed AI approaches in the belief that the full potentials of essentially 'linguistics'-oriented models have yet to be demonstrated: the Eurotra project of the CEC is one example (sect.12); another is the Rosetta project at Philips (sect.17).

In discussion of this issue, it is important to remember the distinction between implicit knowledge and explicit knowledge. Implicit knowledge is that which is incorporated in the lexical and grammatical information of the system. It is the basic knowledge of the language which is a prerequisite for any 'understanding', and comprises therefore linguistic knowledge common to all of those competent in the language (i.e. morphological, syntactic and semantic ‘competence'), and it encompasses linguistic knowledge specific to the particular sublanguage(s) of a text. Explicit knowledge is the extra-linguistic knowledge which is (or can be) brought to bear in the interpretation and disambiguation when implicit linguistic knowledge is insufficient. Explicit knowledge can in fact be of two kinds: the knowledge of the subject(s) of a text which a reader calls upon (i.e. both general knowledge of reality, facts and events, and a more specific 'expertise'), and the pragmatic knowledge acquired in the course of reading, interpreting and understanding the text itself (i.e. dynamically acquired or learnt knowledge of the facts, events, opinions, suppositions, etc. described in the text). Of course, the boundaries between all these kinds of knowledge are very fluid: what has been learnt from one text may be applied as knowledge in understanding another; sublanguage knowledge is inextricably bound up with subject 'expertise';

and what is 'common' linguistic knowledge changes over time to include what in the past may have been esoteric sublanguage knowledge.

The question is not whether understanding has a role in MT but how large a role it should have, and specifically how far MT systems should go in the direction of programs for natural language understanding (NLU). In general NLU programs (mainly in the context of AI) have been designed for specific purposes or tasks, e.g. in domain-specific retrieval systems, in systems for paraphrasing, and for composing newspaper summaries; thus 'understanding' is effectively determined by specific expectations of text content or potential users' interests, often expressed in terms of 'scripts' or 'schemata'. NLU systems are concerned above all with the content and the message, and not the specific linguistic (discourse) framework in which the content is conveyed; once the 'message' has been extracted, the linguistic 'form' can be disregarded. But MT has to embrace all aspects of texts, only some of which are language-independent (universal) and only some of which involve extra-linguistic 'understanding'. It is argued (e.g. Tsujii 1986) that while interlingual concepts can be, indeed must be, defined in terms of extra-linguistic 'reality', facts and theories (i.e. the terminologies of the natural sciences, of engineering, of medicine, etc.), non-scientific vocabulary (which occurs in all texts as the basic 'linguistic' knowledge) cannot be defined other than intra-linguistically and language-specifically, e.g. as semantic networks within a particular language. In any case, the boundaries between 'scientific' sublanguages and 'non-scientific' common vocabulary are vague, indeterminate and changeable. Therefore, although some knowledge and text understanding is language-independent, much is specific to particular languages. Many processes of interpretation in MT are determined by the specific characteristics of the languages concerned; the choice of a TL lexeme may require information which is not expressed in the SL, e.g. when translating into Japanese from English, the relative status of speakers and hearers - the 'understanding' of the English text does not involve this information at all. For MT 'understanding' is not exclusively language-independent. MT must deal with a multiplicity of knowledge sources and levels of description: linguistic (implicit) data of both SL and TL (both sublanguage and common knowledge), information on lexical and structural differences between SL and TL, and explicit (a priori and dynamic) knowledge of the subject(s) of texts.

For such reasons, many MT researchers believe that MT systems should build upon well-founded and well-tested 'linguistics'-oriented approaches, with extra-linguistic knowledge bases as additional components alongside morphological, lexical, syntactic, semantic and text-grammatical information. The dominant framework for most MT systems under current development is essentially that of the transfer models represented by GETA (sect.13), SUSY (sect.15), Mu (sect.26), METAL (sect.5) and Eurotra (sect.12). These projects are founded upon a solid body of well-tested and efficient methods of morphological and syntactic analysis with modular flexible system architectures permitting progressive incorporation of newer techniques. Not surprisingly, many commercial systems recently marketed or under development have adopted this basic transfer model (particularly in Japan.) However, as experience with sophisticated theoretical modelling of linguistic processes has increased, the popularity of the more ambitious interlingua approach has also grown, e.g. Rosetta (sect.17), DLT (sect.18), Translator (sect.10), ATAMIRI (sect.41), ETL (sect.28), LUTE and LAMP (sect.36).

3. Transfer

What distinguishes all MT systems from natural language understanding systems is the prominence of 'transfer' operations. At the most trivial level this means only that NLU systems are monolingual while MT systems are bilingual or multilingual. At the heart of MT are the conversion of SL lexical items into TL lexical items and the transformation of SL structures into TL structures. MT systems differ in the size and relationships of three basic processes: monolingual analysis of SL texts, bilingual conversion of SL expressions and structures into TL expressions and structures, and monolingual synthesis (or generation) of TL texts. Monolingual processing is language-specific and independent of any other language, bilingual processing is oriented to a pair of languages: the SL and the TL. At the two extremes are 'direct translation' systems and 'interlingua' systems. In 'direct' translation of the word-by-word variety there is no (or scarcely any) monolingual treatment, rather the program consists (almost) entirely of rules converting SL forms into TL forms. In 'interlingua' systems there are two monolingual components: SL analysis results in an intermediary (interlingual) representation which is the input for TL generation; there is no bilingual conversion from SL to TL, only transformation into and out of an intermediary language IL (interlingua). In practice these 'pure' forms of direct and interlingua MT are rare. Most MT systems incorporate some monolingual analysis and synthesis and some bilingual transfer. Where they differ is in the 'depth' of analysis, the 'abstractness' of transfer processes, and the amount and type of semantic and extra-linguistic information. Consequently, one way of classifying MT systems may be in the broad groups: morphological transfer, syntactic transfer, semantico-syntactic transfer, semantic-conceptual transfer.

Most of the early 'direct' MT systems were morphological transfer systems: monolingual analysis was limited generally to the establishment of word-class membership (grammatical categories), e.g. noun, verb, adjective, on the basis of inflections, conjugations and local cooccurrence rules; monolingual generation was likewise limited to the production of the correctly inflected TL forms; the bulk of the programs were devoted to SL-TL transfer processes - lexical translation and word order changes (often prompted by specific lexical items). A modern example of a 'morphological transfer' system is the CADA program for converting texts between closely related languages (sect.9): the approach is valid only where there is a high degree of lexical correspondence, of morphological equivalences, and little (if no) differences of syntactic structures.

Syntactic transfer systems appeared early in MT research (the approach of Yngve at MIT is the best known example). Monolingual analysis concentrates on the establishment of 'surface' syntactic analyses (e.g. phrase structure or dependency trees), without seeking to eliminate structural ambiguities (e.g. the scope of negation, the relationship of a prepositional phrase within the sentence). In general, such syntactic analysis is normally preceded by morphological analysis, but sometimes there is a single morpho-syntactic stage: the decision is largely determined by the nature of the SL. Bilingual transfer involves both SL-TL lexical conversion and transformation of SL syntactic structures into equivalent TL syntactic structures. Monolingual TL generation may then establish the appropriate TL syntactic and morphological forms. Not infrequently, however, TL generation is embraced largely (even wholly) by processes within transfer. Recent examples of (basically) syntactic transfer systems are to be found in many of the Japanese commercial systems.

More common, however, is a semantico-syntactic transfer approach. In such systems, monolingual analysis seeks to eliminate syntactic ambiguities and to provide single representations for synonymous syntactic (sub)trees. An increasingly common method is the establishment of the case roles of nominal expressions (e.g. agent, patient, recipient, instrument, location, etc.) and their relations to verbal (action) expressions. Case roles and relations may be very similar for both SL and TL (indeed it is the assumed or presumed 'universality' of case relations which makes the approach so attractive). Transfer processes, therefore, may involve little change of SL structures into equivalent TL case structures; the burden of transfer is on SL-TL lexical conversion and any changes of structure resulting from differences of TL valencies. Monolingual generation involves the production of TL syntactic and morphological forms from the TL case-structure representations output by the transfer phase. This MT type has been found particularly appropriate for Japanese, with its relatively free word order; and many Japanese projects may be classed in this group.

A further level of abstraction is reached in what may be called 'semantic-conceptual transfer'. Monolingual analysis might include the determination of pronominal antecedents, the scope of coordination and negation, the logical relations of predicate and arguments, the textual functions of theme, rheme, and information presentation. The analysis procedures might be based on semantic features and semantic networks (hyponymic and thesaural relations), on knowledge databases and inference mechanisms; both linguistic (language-specific and universal) and extra-linguistic (language-independent) information may be invoked. The SL interface representations may combine 'linguistic' semantic (both language-specific and universal) and non-linguistic conceptual (language-independent) elements. They may well transcend sentence boundaries and embrace paragraphs. Transfer is then concerned with the conversion (perhaps minimal) of SL-specific lexico-conceptual elements into TL-oriented elements, and perhaps not at all with structural changes (since these may be interlingual). The theoretical distinctions between semantic and conceptual information are in practice blurred, although some MT systems concentrate exclusively on 'linguistic' processes (implicit knowledge) and make little use of extra-linguistic (explicit) knowledge. By contrast other AI-inspired models have been more 'conceptually' oriented (e.g. the conceptual dependency representations of Schank and his colleagues). At this higher level of abstraction or universality, the distinction between transfer systems and interlingua systems becomes less valid; all that remains of transfer is the 'adjustment' of a SL conceptual framework (determined by non-universal lexical and structural parameters) to a TL conceptual framework.

There are two basic approaches to interlingual representations. The first is to extend the progressive abstraction of SL and TL interface elements so that all elements (lexical as well as structural) are language-independent, interlingual and perhaps 'universal'. (The goal of nearly all interlingua MT projects is multilingual translation; while, in principle, interlinguality may mean only shared neutrality for a particular SL-TL pair, it usually implies a presumed universality.). In the development of interlingual representations, the (assumed) universality of propositional and intensional logic has been particularly influential; logical formalisms have long constituted essential features of semantic representations (e.g. predicate and argument structures) in both linguistic theory and computational linguistics, and it has been further buttressed by the AI use of inferencing in language 'understanding'. This conception of a conceptual (semantic) and logical interlingua is to be found in various forms in many current projects, e.g. the Translator project, the LUTE project and the Rosetta project. The other approach to intermediary representations is to take an existing 'interlingua' (IL) - an international auxiliary language or a 'regular' natural language - and to devise procedures for converting SL representations into IL representations (texts) and from IL representations into TL texts. The first option has been taken by DLT, which has adopted Esperanto as its interlingua; the second is to be seen in the ATAMIRI project, where the interlingua is a South American language Aymara (sect.41).

There is general agreement that at the semantic levels (whether in transfer or in interlingua systems) interface structures must retain lower-level information if satisfactory coherent TL texts are to be produced. Interfaces are thus multi-level, combining syntactic categories, phrase structure relations, dependencies, case roles, thematic functions. Examples are to be found in the transfer systems GETA and Eurotra, and in the interlingua systems DLT and LUTE.

Recent developments in the variety and complexity of transfer and interface structures have highlighted the theoretical issues of transfer (e.g. Hauenschild 1986, Luckhardt 1987a, Somers 1987a, Tsujii 1987). Transfer was of course always recognised as important, but in the past the problems of NL analysis were seen as the more urgent, as the more intractable, and as the principal impediments to good quality MT output. Now transfer and generation are considered of nearly equal importance. Generation has received much attention in the AI field, because without good text production man-machine communication is deficient (e.g. in question-answering and information retrieval systems). It is also beginning to receive more attention in the MT context (McDonald 1987, Nagao & Tsujii 1986). However, it is transfer that distinguishes MT from monolingual NL processing in AI, and in most of computational linguistics - and it is transfer which must also represent a central feature of any 'theory' of translation. A particularly valuable contribution is the monograph by Luckhardt (1987a), who discusses the relationship between analysis and transfer (in general the more analysis the less transfer, except that deeper analysis does not always simplify transfer or generation), the difference between explicit and implicit transfer, the types of unit transferred (lexical, structural, and categorial), the importance of valency and case theories as the basis of transfer interfaces, and who has thereby laid the foundation for future research on transfer.

Recent contributions have indicated other areas of general agreement. One is that in bilingual systems, e.g. the ENGSPAN system, transfer and interfaces can be at a relatively shallow level, while in multilingual systems it must inevitably approach 'semantic-conceptual' levels of analysis, as in Eurotra, or the systems must adopt an interlingua approach, as in DLT. The context here is that of systems designed from the beginning as multilingual systems, not the adaptation of bilingual systems to new target languages (as in Systran, sect.1 below). Another is that there may well be three aspects (perhaps stages) of transfer: the conversion of SL-oriented structures into SL 'neutral' interfaces (in which the structures peculiar to the SL are reduced), the conversion to TL 'neutral' interfaces, and the conversion to TL-oriented structures. The first corresponds in the Mu system (sect.26) to the 'pre-transfer loop' and the third a 'post-transfer loop'. Similar distinctions are made by Schubert (1987) in the 'metataxis' procedures for the DLT system. A third observation (Boitet 1987) is that in multilingual systems involving cognate languages it may well be feasible to establish common interfaces (e.g. a common interface for the Romance languages, or a common interface for Scandinavian languages), and thus reduce the complexity and number of transfer components. The increasing interest in case grammar as the basis for transfer systems (particularly in Japan) has demonstrated that traditional case and valency analyses are often unsuitable or insufficient and have to be changed or adapted for MT purposes - as they have, e.g., in Eurotra (cf. Schmidt 1986). In particular Somers (1986, 1987b) argues that valency analysis needs to be extended to nominal structures. Finally, the need for adequate treatment of discourse relations has become increasingly pressing. A number of researchers are investigating the place of text grammars in MT environments, e.g. at Grenoble (sect.13), at Stuttgart and Saarbrücken (sect.16 below), in the Translator project (sect.10) and in the research at ETL (sect.28).

In a sense all MT approaches have some merit because they are all attacking some aspect of the translation problem; all projects encounter the same issues in the end: semantic relativity, structural mismatch, no solid language universals, the complexity of dictionaries, the problem of 'ill-formed' texts and of metaphor, of dynamic change. The difficulty lies in finding an overall strategy which can encompass most of the facilities required. At present the transfer approach still seems most suited. As Tsujii (1986) points out it does not exclude interlingual elements nor 'understanding' and inferencing, rather it encourages research on the precise role and contribution of these features to genuine translation.

4. Grammars and programming

As regards the computational aspects of MT there is a large measure of convergence on what are called 'unification grammar' formalisms (Kay 1984, Shieber 1986) and on non-transformational grammatical theories, particularly Lexical Functional Grammar (Kaplan & Bresnan 1983), Generalized Phrase Structure Grammar (Gazdar et al. 1985) and Montague grammar (Dowty et al.1981). Implementation of LFG is to be seen in many MT projects, nor only experimental ones but even in the commercial development of Weidner systems (sect.6); GPSG is favoured increasingly in experimental systems, e.g. NASEV (16) and NTRANS (20); and Montague grammar has been implemented not only in Rosetta (sect.17) but also in LUTE (36). At present the most appropriate programming environment for parser and generators based on such formalisms is considered to be Prolog, which has begun to replace Lisp as the favoured language in many NLP situations (the relative merits of Lisp and Prolog for MT are discussed by Loomis 1987). Although Prolog was adopted as the basic programming language for the Japanese Fifth Generation project, it is to be found in relatively fewer MT projects in Japan than other languages - C is favoured increasingly in systems designed for commercial development.

Increasing attention has been paid to creating linguistic software and appropriate workstations for linguists to develop grammars, parsers and generators. Examples are the software environments developed at Grenoble (sect.13), at Saarbrücken (the SAFRAN workstation, sect.15; cf. Licher et al. 1987), at Utrecht for DLT (sect.18), at Kyoto (sect.26: the GRADE language), and for the METAL system (sect.5; cf. White 1987).

There is no pretension that the basic future framework has been decided. Among the many problems for future MT research (Tsujii 1986) are the content and structure of multi-layered representations, the integration of different levels of 'understanding', the development of robust (fail-safe) parsing of incomplete (or 'ungrammatical') texts and the treatment of weak semantic constraints (preference semantics and metaphorical usage), the development of robust and flexible frameworks for new developments, the efficient application and appropriate formalization of text linguistics, the problems of large and complex semantic networks, and the practical difficulties of ensuring consistent coding, particularly the lexical coding of large dictionaries.

5. Further developments

The emphasis of most MT research is towards systems which produce written translations and which start from complete finished texts. In recent years, research has expanded into new types of translation. There has been considerable interest in the prospects of telephone translation, a major project in Japan involving by governmental sponsorship (the ATR investigation, sect.28) and a project by British Telecom in the United Kingdom (sect.20). There are also signs of interest by Japanese companies, e.g. Toshiba (sect.32). There are few hopes of working systems in the near future; 15 years before a prototype is one of the more optimistic forecasts.

One new arena for MT offers more immediate prospects of working systems. It is recognised that AI knowledge-based systems produce 'paraphrases' rather than translations, i.e. they concentrate on the essential 'message' and disregard surface expressions. There are already programs for checking scripts during composition (e.g. the 'expert editor' of Smart, sect.3) and allowing interactive rewriting for reducing the complexities of MT operations. There is the prospect of systems which combine composition in the user's own language and simultaneous translation into another. The most immediate application would be conventional business correspondence, but extension to other spheres would surely follow, particularly if there were some speech input and output. The research at CMU, sect.10, and at UMIST, sect.20, are first steps in this direction.

A more distant prospect must be systems combining translation and summarization. The idea of producing summaries of foreign language documents for administrators, businessmen and scientists in their own language is almost certainly more attractive than rough translations of full texts. AI researchers and others have conducted small scale experiments on summarization in restricted domains, but it is already apparent that the complexities of the task are at least equal to those of MT itself.

Greater success is likely with the integration of MT and information retrieval systems. Sigurdson & Greatrex (1987) at the Research Policy Institute of the University of Lund have demonstrated that the technical means are already available for businessmen and scientists to access Japanese databases and obtain automatic translations of abstracts. A number of MT systems exist already which were designed for translating titles and abstracts (e.g. TITRAN and TITUS) and Systran has also been applied to patents (sect.1). In the Japan-Info project of the EEC users will be able to ask for translations of Japanese abstracts (mainly of research reports and other 'grey literature'); they will be produced in Japan using Systran Japan and the Fujitsu ATLAS systems (sect.29 and 30), transmitted to Europe and then to the requesters (Sigurdson & Greatrex 1987). The next step must bean integrated system - this is the aim of the MARIS project at present under development at Saarbrücken based on the SUSY translator (sect.15; Zimmermann et al.1987).

6. Operational systems

There are other examples of research groups looking for direct applications of their experimental work. They include the involvement of the Grenoble group in the French national project (Calliope), the development of the METAL system as a commercial German-English system by Siemens, and the second phase of the Mu project in Japan. Nevertheless, most operational and commercial systems have originated not from academic research groups but from independent companies.

The longest established, and still the leader for 'batch processed' MT, is Systran (sect.1). Originally devised for Russian-English, then on behalf of the European Communities for English-French, Systran is now in worldwide use in an impressive range of language pairs. The recent coordination of separate developments of Systran version under the general ownership of Gachot and the recent online availability of Systran translations should mean the continued vigour of Systran for many years to come. Other 'batch' systems do not have the same international position. PAHO's systems for English and Spanish were created for in-house use only (sect.2); Smart concentrates on large systems tailored to particular organizational needs, so far exclusively in North America (sect.3); Logos has been successful in the German market but has so far made little impact elsewhere (sect.4); and METAL, also for German-English translation and probably the most advanced 'batch' system, is as yet only at the stage of trial implementations (sect.5).

In the sphere of 'assisted' systems the current leader is undoubtedly WCC (previously Weidner), now owned by Bravis of Japan (sect.6). It now offers an impressively wide range of languages for both VAX and IBM PC equipment, with particularly large sales of its microcomputer-based systems in Japan. Its main rival remains ALPS, which however has recently concentrated more on machine aids for translators rather than interactive systems, and has expanded its commercial base by purchases of translation agencies and bureaux. Other microcomputer 'interactive' MT systems have appeared in Europe and North America more recently, e.g. TII, Tovna and Socatra (sect.8), but undoubtedly the greatest impact has come from Japan.

In the past five years there have appeared on the Japanese market systems for English-Japanese and Japanese-English translation from many of the large electronics and computer companies. Fujitsu and Hitachi were first with their ATLAS and HICATS systems respectively (sect.30 and 31). They have been followed closely by Toshiba (32), Oki (33), NEC (34), Mitsubishi, Matsushita, Ricoh, Sanyo and Sharp (35). In design some of these are ambitious 'interlingua' systems, but most are relatively low-level transfer systems. Nearly all require in practical implementation a considerable pre-editing effort if satisfactory results are to be achieved: in large part, pre-editing is conditioned by the particular difficulties of Japanese (three scripts, lack of word boundaries, high degree of ellipsis, complex sentence structures). However, Japanese operators are accustomed to similar requirements when using Japanese word processors with no translation envisaged, so the extra costs are acceptable.

The range and diversity of MT products now available reflects the recognition of a large potential market. There is undoubtedly more translation work in technical, economic and scientific fields than can be dealt with by competent and qualified translators. An indication of the commercial expansion of the MT field are a survey of translators' work practices by DEC (Smith 1987), a large-scale evaluation of the potential market by Johnson (1985) and a technical evaluation by Balfour (1986). MT and machine-aided translation is now capable of offering at greater speed and at usually lower costs than human translation a diversity of translation products, which did not previously exist, and which can satisfy a wide range of requirements from rough unrevised MT output (for information purposes only) to fully post-edited high quality versions.

However, the introduction of MT in any operational situation always demands considerable effort in compiling special dictionaries for local needs, and their compilation requires appropriate translation expertise - which purchasers of MT systems are not always aware of. The practical problems of integrating MT in translation services are not trivial, but they can be overcome, as illustrated by the experience at CEC with Systran (sect.1); numerous other examples are to be found in papers given at the Translating and the Computer conferences (Picken 1987, 1988), at the Japanese MT Summit (1987) and in the pages of *Language Monthly*; in nearly all instances, companies report higher throughput and greater consistency, but not always lower costs.

It is possible that in future the greatest expansion of MT will come in the provision of rough translations for those with sufficient subject knowledge to be able to overlook the grammatical and stylistic 'mistakes' of present MT systems. These are translations which would not have been done at all without MT, and in this respect MT will be fulfilling urgent needs. Nevertheless, there are dangers; the limitations of MT may not be appreciated by those who are ignorant of translation or of languages. As long as systems are bought and operated by translators, translation bureaux and by companies with experience of translation, the recipients of unrevised MT output are likely to be made fully aware of the limitations of MT versions. The danger will come when, as we may expect, systems require less post-editing, are less restricted in subject coverage, and are purchased by non-translators and those with no foreign language knowledge. It is the duty of the translation and MT community to ensure that the general public is not misled by unrealistic claims and promises.

Developments in the past five years have been more rapid than at any time since the 1960's; MT activity is growing not just in the developed countries of North America, Europe and Japan, but in less technologically advanced countries of Asia and South America. The internationalisation of MT research and of MT implementations is attributable to many factors: the trans-national commercial dimension of translation itself (Systran, Bravis/WCC, ALPS, Logos), explicitly multinational projects (Eurotra project, Japanese ODA project), international collaborative research (e.g. initiated by GETA, Siemens, Fujitsu, etc.), multinational companies (Xerox, IBM), the wider availability of increasingly powerful computer equipment, and the solid theoretical and practical achievements of natural language processing. The vigour of MT is reflected in its mixture of experimental speculative model building and of practically oriented development of operational systems. Machine translation is no longer a slightly suspect academic pursuit (as it was until the mid-1970's in many respects), it has established itself as an important branch of applied science (computational linguistics and artificial intelligence) and as a technology-based industry of international dimensions.

II. Survey of current and recent projects and systems

This survey of recent activity in MT research and production systems is intended to be as comprehensive and accurate as possible; it includes small scale experiments, large projects, government-funded projects and commercial systems, each with brief notes of recent developments as a guide to the extent of current activity. It represents a provisional supplement to the survey published in 1986 (Hutchins 1986), which covered MT research and production systems to the end of 1984. It includes therefore notices of systems already established in 1984 as well as new (or not previously noted) projects. Further information (or corrections) about any MT activity, whether mentioned or not in this survey, will be welcomed. The bibliography at the end of this paper, although substantial, is not designed to be comprehensive but to guide readers to the most recent (and usually most accessible) general accounts of particular systems. (The abbreviations LM and LT refer to the journals *Language Monthly* and *Language Technology*, both of which carry regular items on MT.) The arrangement is broadly 'geographical', starting with systems originating or based in North America (nos.1-11, including Systran, PAHO, Smart, LOGOS, METAL, WCC, ALPS, etc.), passing on to European activity (nos.12-24, covering Eurotra, GETA, Saarbrücken, Rosetta, DLT, etc.), then to the numerous Japanese systems and projects (nos.25-37), and ending with the rest of Asia and with South America (nos.38-41).

1.

Systran is the oldest and the most widely used MT system. (Basic descriptions of the Systran design can be found in Hutchins 1986: 210-215). Ownership has now passed completely out of the hands of its original designer, Peter Toma (who has now set up a private university (Aorangi University) in Dunedin, South Island, New Zealand, in order to promote international conflict resolution) and almost wholly into the Gachot S.A. company. After a complex series of negotiations spread over a number of years, the Gachot company (Jean Gachot and Denis Gachot) has now united all US and European companies with Systran interests. The process had already begun with the acquisition by Jean Gachot of the Systran Institut GmbH and the World Translation Center (WTC) in La Jolla, California, and by Denis Gachot of Latsec, the US branch mainly concerned with military applications. Now the only company with Systran rights which remains outside is the IONA company (headed by Sadao Kawasaki) which owns the Systran Corporation of Japan (cf. 29 below) and the rights to the Japanese programs (Joscelyne 1988).

All the main organisations and users of Systran were brought together in February 1986 at the 'World Systran Conference' organised by the Commission of the European Communities and held in Luxembourg. The proceedings have been printed in special number of *Terminologie et Traduction* 1986, no.1.

The longest standing user of a Systran system is the USAF Foreign Technology Division at Dayton, Ohio, where the Russian-English version has been in use for information scanning since 1970. By 1987 nearly 100,000 pages each year were being translated (Pigott 1988). The quality of raw output can be judged by the fact that only 20% of texts are edited (by the EDITSYS program) - this is the output which has been automatically flagged by the Systran program for not-found words, acronyms, potential rearrangement, potentially suspect adjective-noun and noun-noun compounds, uncertainty in disambiguation, and known problem words (Bostad 1986). The Russian-English system has recently been successfully adapted to the translation of Soviet patent abstracts (Bostad 1985). In addition, the USAF has now introduced German-English and French-English versions, and is developing Italian, Portuguese and Spanish as source languages (Pigott 88); there is apparently also a Japanese version under development (LT4 Dec 87).

Use of Systran at the Commission of the European Communities (CEC) began in 1976 with development of the English-French version, followed soon by French-English and English-Italian, and a pilot 'production service' for these language pairs in 1981. Since then versions have been also developed at Luxembourg for English to German, Dutch, Spanish and Portuguese as well as from French into German and Dutch. Experience so far is that the best quality is achieved for translations into Romance languages and into English, the quality is lower for translation into German and Dutch, where there are particular problems of word order. In the near future, work is expected to begin on German into English and into French (Pigott 1988). Although the reception of MT output has steadily improved (helped by OCR input and links with a variety of word processors), usage is still relatively low - under 2% of CEC translations were post-edited Systran versions in 1987. Nevertheless, future growth is expected with more translators opting to use raw outputs as aids to quality manual translation (i.e. the use of MT versions as pre-translations), and with more recipients of information documents satisfied with rapidly post-edited texts (Pigott 1988).

Other long standing users are General Motors of Canada, where an English-French system produces literature for the Canadian market, and Xerox, where technical manuals written in a restricted English are translated at the rate of 50,000 pages a year into five TLs: French, Spanish and Italian. Other users include the NATO headquarters in Brussels, the Dornier company in Germany, the German national railways (Bundesbahn), the Nuclear Research Center (Kernforschungszentrum) in Karlsruhe, and the International Atomic Energy Authority. A major user is the French company Aérospatiale, which after an initial trial with the CEC systems at Luxembourg, contracted directly with Gachot for use of English-French and French-English systems to translate aviation manuals. Aérospatiale hope that 50-60% of their needs will be met by unedited Systran output; Habermann (1986) has reported that researchers at the Kernforschungszentrum are very satisfied with the raw output from the French-English version.

In the last 5 years a number of bureau have started to offer clients translation services using Systran (Pigott 1988). Gachot in Paris is one of them; others include ECAT (European Centre for Automatic Translation, Luxembourg), the Mendez service bureau (in Brussels) and CSATA (Italy). Recently, Systran has been offered by ECAT to companies on Esprit projects as an on-line translation service (Siebenaler 1986).

The most striking development which Gachot has introduced has been to make Systran (and some of the Systran dictionaries) accessible to the 4.5 million users of the French Minitel network (cf.14 below). Beginning with English-French, Gachot has now added Spanish-English, English-Spanish and English-Portuguese. The implications may well be revolutionary. Until now MT has been primarily for academics and for professional translators; Gachot has brought MT within reach of the general public. Its most effective use is likely to be for international business messages. The use by French students to help them with their homework is more dubious.

While the ownership of Systran rights was divided, systems developed in a relatively uncoordinated fashion. Development of Systran systems was undertaken with joint contracts between major users and the Systran companies such as (in the CEC case) the World Translation Center, World Translation Company of Canada, the Franklin Institut and the Systran Institut, and Informalux (Luxembourg). In most instances, the users developed the dictionaries, while the Systran companies worked on the software modules. Although WTC and Latsec, therefore, maintained a guiding role, there were inevitable divergences as long as dictionary construction was adapted to specific purposes, at USAF, Xerox, General Motors and within the CEC. Systran is essentially a lexicon-driven system (in common with most MT systems of the first generation) and inevitably the diverging 'philosophies' of dictionary construction resulted in diverging system types. The dangers were always recognised. Now, in bringing together Systran development in one organisation, Gachot can ensure that the coordination of divergent programs will continue more vigorously, with the aim being a total unification of all Systran systems. At the same time, Gachot benefits from the large dictionary databases which have been established by the CEC, Aérospatiale, etc.

The increasing modularity of the Systran designs will aid convergence. At the CEC the analysis programs for English and French have been adapted relatively easily to new target languages. The experience at Xerox with a single SL program (for English) and multiple TL programs is further evidence of the potential of Systran. According to Ryan (1987) there is now a 'common trunk' of procedures for the Romance languages, and the German and Dutch TL modules are also common to more than one pair. Just as the English SL modules are transportable to new targets, the French SL components are being adapted at present for multi-target systems. All this does not mean that Systran is a 'multilingual' system, since each version has to be specially designed for a particular language pair in one direction; but it does mean that the addition of a new version is becoming progressively easier. Gachot has been reported as developing a number of new versions: English-Italian, English-German, German-English, and to be working on Arabic, Portuguese and Spanish. (The problems of Arabic have been outlined by Trabulsi 1986.) In total, "Systran now offers 15 operational language pairs. These include English into 8 languages: French, German, Italian, Spanish, Portuguese, Russian, Japanese and Dutch; French into English, German and Dutch; and Russian, German, Japanese and Spanish into English. English-Arabic is under development, while pilot systems exist for 6 other language pairs: German into French, Spanish and Italian; and Chinese, Portuguese and Italian into English." (Ryan 1987)

There are now very large Systran dictionaries: USAF Russian-English has over 200,000 single words and over 200,000 expressions; CEC dictionaries for 4 languages have between 100,000 and 200,000 words; for the English-Japanese system it is planned to have 250,000 word dictionary of scientific and technical vocabulary and over 200,000 of medical terms. It is Gachot's aim to achieve by 1990 a quality level averaging over 96% for all the 12 language pairs currently available, although it is recognised that raising the quality of English-German and English-Russian, currently estimated at 67%, will be a major achievement in this time span.

After a period of threatening fragmentation into different 'dialects', Systran is entering a phase of consolidation guided by the Gachot interest in developing further what has for many years been the most widely used and undoubtedly most successful mainframe MT system.

2.

At the Pan American Health Organization (PAHO) there has been continued development of the Spanish-English (SPANAM) and English-Spanish (ENGSPAN) systems designed to translate a wide range of subjects and of document types in the broad field of medicine and health. The PAHO systems are well known as robust empirical systems based on well-tested MT techniques. The aim was practicality not experimentation. SPANAM became operational in 1980, ENGSPAN in 1985; the two systems are described in detail by Vasconcellos & Leon (1985/1988), and summarised by Hutchins (1986: 220-222).

Most research development has concentrated on ENGSPAN (Leon & Schwartz 1986). This is a batch-processing system with post-editing, but with no pre-editing and no restrictions on content or style. It is based on a transfer design, limited primarily to lexical and syntactic transfer, and with minimal semantic analysis. Morphological and syntactic analysis is by an ATN parser, written in PL/I. There are proposals to convert the PL/I code to the C language for running on a microcomputer.

SPANAM, fundamentally a 'direct translation' MT system, has undergone no basic changes since it first became operational, but there are proposals to incorporate features of the more advanced ENGSPAN in some future developments. It is now well-established at PAHO and for its on-screen post-editing special facilities have been developed, which are described by Vasconcellos (1986, 1987). The aim has been to devise functions ('macros') for frequently recurring actions: replacements, deletions and insertions (e.g. of articles), inversions (e.g. N of N $\rightarrow$ N N), etc. For the treatment of phrase structure changes it is argued that as far as possible the informational sequence (theme-rheme articulation and ideational presentation) of the original Spanish is to be preserved in the English version. For example, the 'raw' output:

For its execution there has been considered two stages...

could be changed to:

Its execution has been conceived in two stages...

As in this case this may mean that syntactic functions have to be altered, but it is often easier to do this than to shift around large segments of text. Undoubtedly such facilities would assist post-editing work in other practical MT operations.

3.

The unique feature of SPANAM and ENGSPAN is that they were designed for, researched by and developed within and for the sole use of a single organisation. Whether this approach will ever be repeated is unlikely. In the future, the more probable course for organisations like PAHO will be to call in a company like that set up by John Smart in New York.

Smart Communications Inc. provides two basic products, the Smart Expert Editor and the Smart Translator (Mann 1987). The editor (MAX) is a batch-oriented text analyzer using a rule-based expert system and a specialized terminology knowledge base. The rule base contains 2500 generalized grammar and syntax rules for technical writing, the knowledge base includes information specific to the needs of a particular company (e.g. on dangerous chemicals). MAX acts like good copy editor, producing a report for the technical writer to act upon. As a by-product the result is easier to translate.

The Smart Translator thus operates on a restricted grammar and lexicon of a source language. It does not attempt to deal with ambiguities and vagueness, which are to be eliminated by the Editor. Smart's restricted language was based initially on Caterpillar English, devised for the Caterpillar company and subsequently modified and expanded by Xerox as their Multinational Customized English for use with the Systran system (sect. 1 above).

Smart Communications therefore provides tools for economical on-line help in writing clear, safe documentation and for translating it into several languages. It has been active since 1972, and there are now over 30 companies using SMART software, including Citicorp, Chase, Ford, and General Electric. Smart's largest customer to date is the Canadian Ministry of Employment and Immigration. A system was produced to translate job vacancies from English into French and vice versa. In operation since 1982, job descriptions are input at 5000 terminals across Canada, over 100,000 a year, and translated at a rate of up to 200 characters per second. The aim is not perfect translation; 90% accuracy is acceptable, since there are bilingual secretaries who can post-edit. The officials worked closely with Smart in tailoring the system to their particular requirements and they continue to maintain the knowledge base and update the dictionaries.

The Smart Translator has been implemented from English into French, Spanish, Portuguese, and Italian and vice versa. Some work has been done on German and on Japanese, and there are reports of future plans for English into Greek and Turkish (Rolling 1987b). There are no plans to implement cross-language pairs, e.g. French-Spanish, but Smart is investigating a French editor.

At present there are no customers outside North America, although there seem to be plans to open a base in England shortly. Like others, Smart has ideas also for a business letter writing 'kit' which would enable businessmen to write letters in any language by selecting parts needed. There might also be development of a more active interface, i.e. in effect an interactive MAT system, but Smart is doubtful that users really want this kind of system.

4.

The LOGOS English-German system was demonstrated in 1984 and joined the German-English version (first available in 1982) on the market, mainly in West Germany (largest user is the computer firm Nixdorf). Both systems run on a Wang OIS140 with a minimum of 80MB hard disk; and are now also available in versions for IBM (VM/CMS) mainframes. LOGOS has European offices in Frankfurt and Zurich, with the former providing a bureau service since 1985. (Recent details of LOGOS are given by Wheeler 1985.)

Operation involves a preliminary run through the source text for missing vocabulary, the entry of new lexical items by the user (prompted by the computer for information on syntactic categories, grammatical case, inflection type, and semantic codes). The text is then translated in batch mode: morphological analysis, syntactic analysis, treatment of nominal subtrees, check for idioms, lexical transfer (guided by semantic codes), identification of subjects, objects, etc., and generation of target text - transfer and generation are combined (in a series of stages) in the fashion of some earlier 'direct translation' systems, i.e. LOGOS can be described now as a 'hybrid' syntax-oriented transfer system (Hutchins 1986: 255-257). Output is then post-edited using the Wang word processing facilities.

LOGOS provide customers with a basic bilingual dictionary of over 100,000, to which they can add their own specialised terminology; however, customers' additions cannot include verb entries, since these demand complex coding and have implications for the efficiency of the system as a whole. Software updates are supplied at no extra cost. There are reports of progress with more language pairs: English-French (in Canada), English-Spanish, and German-French for the Walloon administration in Belgium (Rolling 1987).

5.

The METAL system was developed largely for bi-directional English and German translation at the University of Texas. Since 1978 it has been fully supported by the Siemens company in Munich. A commercial prototype is being tested at present, much earlier than Siemens had initially expected when they began sponsorship. A basic description of the system is to be found in Bennett & Slocum (1985/1988), and summarised in Hutchins (1986: 248-254): METAL has a modularized transfer design, with monolingual and bilingual/transfer dictionaries, a bottom-up chart parser, fail-safe heuristics, it operates on a Symbolics 36-series Lisp machine and is designed for batch processing (estimated speed: 200 pages per day) with post-editing on PC workstations, with sight either of both source text and target version together or of final target text only.

METAL will undoubtedly be the most sophisticated transfer system to be commercially available. The German-English version (the oldest) is being used to produce translations in several pilot operations (in Switzerland and elsewhere) and is due for imminent launch (Schneider 1987). METAL has been applied primarily in the fields of data processing and telecommunications. Future customers will benefit substantially in the construction of their specialised dictionaries by access to Siemens' multilingual term bank TEAM.

Other versions also under development involve Dutch, French and Spanish. Research is being sponsored by Siemens in Munich itself, at the University of Texas (Austin, Texas, USA), the University of Leuven (Belgium) and in Barcelona (Spain). An English-Spanish prototype is expected shortly (Schneider 1987), and some progress has been reported on a Dutch-French version (see also sect.19 below). Most advanced appears to be the English-German system; the early stages of the research in Austin (Texas) on this version of METAL is reported by Liu and Liro (1987).

6.

While Systran is the most widely used mainframe MT system, the Chicago-based WCC (World Communications Center, formerly Weidner Communications Corporation) is the market leader in microcomputer interactive systems. Since early 1988 the company has been wholly owned by the Japanese translation house, Bravis (which had a majority share in Weidner since 1984). Other WCC offices and bureaux have been established since mid 1986 in Toronto and in Europe (where it trades under the name Weidner Translation Europe Ltd.)

The WCC interactive systems (described in Hutchins 1986) now run in two versions: the MacroCAT systems on DEC MicroVAX II machines, and the MicroCAT systems on IBM PC/XT (and compatibles). Currently (February 1988), the MicroCAT is available for translation from English into French, German, Italian, Japanese, Portuguese and Spanish, and from French, Japanese and Spanish into English. Bravis has been particularly successful in sales of the Japanese to English systems, with a reported figure of 3000 packages sold in Japan (where it is marketed as MicroPack)

Other languages are being added in the near future. Arabic versions are at present under development, and, as reported by Darke (1986), under pilot testing in Saudi Arabia and to be marketed shortly. A MacroCAT system for English to Norwegian translation (ENTRA) was developed in a collaborative project between WCC and the University of Bergen, using as its starting point the existing English-German software (Brekke and Skarsten 1987). ENTRA is first MT system for a Nordic language (except for the Danish component in Eurotra, and some very early experiments in the 1960's). It has been restricted initially to the language of the petroleum oil industry, with some current expansion of dictionaries to computing technology. The project's particular problems were getting word order right, treatment of English genitive relational phrases, and the problems of complex noun compounds. Preliminary evaluations indicate an acceptable quality for post-editing, but fuller quality testing has yet to be done.

More fundamental research is reported from the WCC research group at Provo, Utah. Significant progress has been made on more advanced software (System II) based on lexical-functional grammar (*Language Monthly* 54, March 1988, p.3). If so, this is confirmation of the desire of commercial companies to improve the quality of their product by exploiting the latest advances in computational linguistics. In this way, WCC systems may progress beyond what they are acknowledged to be, 'computer-aided translation systems', providing primarily facilities for dictionary lookup and text processing and only minimal translation.

7.

By contrast, ALPS (Automated Language Processing Systems Ltd.) appear to have drawn back from any aspirations in the MT market. In recent years ALPS has purchased a number of translation bureaux, notably TTI, and is starting to market medical expert systems (LM 50, Nov 1987). The ALPS aids for translators are now in three packages (for IBM PC AT's): 'Transactive' is the interactive translation program (now available for English into French, German, Italian and Spanish and from French into English); 'Autoterm' is an automatic terminology lookup package available for many languages (English, French, Spanish, Italian, German, Danish, Dutch, Portuguese, Norwegian); and the 'Translation Support System' is a multilingual word processor with facilities for terminology management, source text analysis, terminology frequency analyses, text transfer, and automatic word counting.

ALPS has expanded its text processing tools with three further packages: 'ABC Word', a "writer's assistance program", providing access from the user's terminal to a monolingual dictionary (Miriam Webster), to a thesaurus and to bilingual dictionaries (English into French, German and Spanish); 'PeriPhrase', a "rule-based linguistic software tool" suitable for the development of natural language interfaces, e.g. for computer assisted instruction and compiler development (for Unix and Xenix environments); and a 'Computer Analysis Program' for building tailored glossaries, and devising writing protocols.

It is clear that MAT is now becoming a much less significant component of ALPS activities. The linguistic ingredient in the ALPS aids was always slight, and there is no evidence that research is being pursued on improving the quality of the ALPS translation modules.

8.

Within the last few years a number of rivals to ALPS and WCC have emerged in the area of interactive systems and machine aids. The following are those which have come to notice.

ALPS' glossary management package ('ABC Word') has a rival in LinguaTech's Mercury (distributed in Europe by InfoARBED under the name Termex). This represents the Level One of Melby's proposed translator's workstation (Melby 1985/1987). Mercury provides multilingual glossary construction (at present for English, French, German and Dutch) and telecommunication access to remote term banks. Level Two would provide facilities for analysing texts, producing text related glossaries, etc. (as already found in ALPS Translation Support System and in the INK Text Tools). Level Three would be the interface to a MT system, with facilities for post-editing.

Recently, the TII company (Telecommunications Industries Inc.) launched its 'translating word processor', the TWP/70, designed to run on IBM PC-compatible microcomputers with 768K RAM and a 30MB hard disk. The available software provides bi-directional translation for English-Spanish, English-Russian, English-French, shortly to be extended by software for English from and into Italian, German and Portuguese. The system is apparently an interactive system (with split-screen dual language display), and designed expressly as an 'aid' for translators.

The TOVNA Ltd. (originally Israel-based, now in London) have announced a system for Sun-level workstations (LT5 Jan 1988). The system (Tovna is a Hebrew homonym for 'software' and for 'insight') is claimed to be a 'learning system', augmenting and amending linguistic rules by AI inferencing, giving capacity for 'commercial quality' translation between any pair of languages: English-French is to be released in March 1988, to be followed by German, Spanish, Russian, and with plans for Arabic.

Linguistic Products (of Houston Texas) is a company set up by Ralph Dessau and George Mallard to market the interactive CAT packages they have developed for English and French, English and Spanish, English and Swedish and English and Danish. Their first product, an interactive stand-alone system for Spanish-English was launched in February 1985. It is apparently a straight word-for-word direct translation system, with differences of word order dealt with by a syntax subroutine after lexical transfer and generation of TL word forms. Customers receive a 70,000 word dictionary, which includes all verb forms (i.e. there is no morphological analysis), but there are facilities for compiling personal word lists. The designers admit the limitations of the simple system, and the fact that it cannot handle complex text. Its selling point is cheapness (at $480 operating on IBM XT microcomputers); 80 copies of the Spanish-English system have been sold, and some in use by officers of the Customs and Excise in Texas. By 1988 the company expects to have 16 two-way language pairs on the market (LT4 December 1987).

In Canada, the SOCATRA company of Montreal (Société Canadienne de Traduction Assistée), founded in 1981 by Claude Richaud, has developed a machine-aided system XLT for English-French translation and are offering to undertake translation for clients on a commercial basis. The system incorporates an ATN parser, context-free grammars, some semantic analysis and the use of 'fuzzy logic'. Text can be input by OCR, from diskettes and by telephone. Facilities for customized glossaries are provided. Clients can choose to receive output either unrevised, or revised by SOCATRA's own translators. Translation speeds of 60,000 words per hour are claimed. (SOCATRA brochure 1988).

9.

One consequence of increased awareness of MT and computational linguistic techniques is the development of limited systems for specific well-defined aims.

One example is the (apparently) one-off program, TRANSOFT, developed at Johns Hopkins University (Moore et al. 1986) to provide a draft translation of a German textbook. All the words were entered in a bilingual dictionary and assigned simple syntactic/semantic categories. The first stage of translation involved the rearrangement of German sentence structures into English-like structures - for this a set of 'parsing tables' were recursively applied. The second stage involved the replacement of German lexical items by English words. In brief TRANSOFT was a simple 'direct translation' system for word-by-word translation devised for one specific task. It is possible that further exercises of this nature will become common in the future.

A much more significant example, and one which may have implications for other MT projects, is the Computer Assisted Dialect Adaptation (CADA) approach to translation between closely related dialects, which has been under development since 1979 by David Weber (Summer Institute of Linguistics) and others in South America. The method, it is stressed, is designed for members of language families which are closely related lexically, syntactically and semantically. It is not seen as a MT system but as a computer aid for translation within well defined situations. The method focuses on systematic differences, aiming to account for 80% of translation work between the dialects. Residual problems are left to human editors who also make any stylistic changes. Programs are written in C programming language and operate on DEC microcomputers. There are the following stages: input (strips information not required: capitalisation, punctuation, non-alphabetic characters), morphological analysis (decomposition into roots, suffixes, dealing with morphemic variation; indication of grammatical functions of suffixes), lexical transfer (using bilingual root dictionary), synthesis (using TL morphemes and information about suffix functions). In overall strategy the method is of the 'direct translation' type operating essentially at the morphological level only; it is valid primarily for translation between dialects of agglutinative languages. A description of CADA for dialects of the Tucanoan family is given by Reed (1985), Kasper & Weber (1986) and Barnes (1987).

The designers and users emphasise the amount of linguistic preparatory investigation demanded if CADA is to work well. It has in fact been used widely in South America, for languages in Peru (Quechua, Campa), Colombia (Tucanoan), Guatemala (Cakchiquel), and is now being developed in the Philippines, Ecuador (Quichua), Brazil (Tupi). Whether the method can be adapted for language families where more syntactic analysis is necessary is still an open question. Nevertheless, it has been suggested that the approach might also be valid for some Romance language pairs (e.g. Spanish and Portuguese) and perhaps between Scandinavian languages.

10.

In recent years there have been a growing number of experimental MT projects in the United States and Canada. Many of them have been inspired to exploit the power of new programming languages (such as Prolog), to experiment with recent advances in formal linguistics (e.g. lexical-functional and unification grammars) and to apply techniques and methods of Artificial Intelligence in Knowledge Based MT systems.

One of the earliest experiments of this kind has been the project at Colgate University on the interlingua system TRANSLATOR (Nirenburg et al. 1986, 1987, and summarised in Hutchins 1986: 282-283). Extra-linguistic knowledge is incorporated in SL and TL dictionaries, and is utilized by the 'Inspector' module during the transfer stage to assist disambiguation. A major distinctive feature of TRANSLATOR is the inclusion of discourse information in the interlingua representation (the 'IL text' representation). Recent developments in the treatment of discourse phenomena are reported by Nirenburg and Carbonell (1987). They describe the means adopted for recording information about co-reference and topic-comment structures. The purpose of the 'IL text' component is to combine (a) 'meaning representations' of the text derived from semantic relations of the text itself and extra-textual inferences and enrichments derived from a 'knowledge base', and (b) structures relating tokens of the IL lexicon which reflect the discourse presentation (including cohesion markers, modal context) of the SL original. While the eventual aim is a fully automatic system, the designers intend to incorporate interactive facilities for analysis and disambiguation in order to augment the knowledge bases dynamically.

At Carnegie-Mellon University (CMU) research continues on the knowledge-based MT system designed by Tomita and Carbonell (1986; also Carbonell & Tomita 1987). The project intends to capitalise on the advantages of the entity-oriented approach of Hayes (1984) for the representation of language-independent conceptual knowledge in a specific domain (in the case of CMU the domain is doctor-patient communications). The formalism of Functional Unification Grammar (Kay 1984) will be employed to represent language-specific (but domain-independent) information about the grammars of the languages involved (initially English and Japanese, but later Spanish, French, German and Italian). From these two knowledge bases will be compiled automatically at the time of translation a single large LISP program for a "very efficient real-time parser".

The other MT research at CMU is that of Tomita (1984, 1986), who has been concentrating on the best methods of conducting interactive dialogues for MT. Part of the answer is the establishment of suitable general-purpose templates, e.g. to disambiguate a PP relation (whether linked to preceding NP or VP) the templates might be: "The action (VP) takes place (PP)" and "(NP) is (PP)":

I saw a man in the park.

1. The action (*saw a man*) takes place (*in the park*)
2. (*the man*) is (*in the park*)

The other part of the solution is to delay asking disambiguation until all the possible parses, which have not so far been deleted, are available together. However the procedure is not without considerable problems as Whitelock et al (1986) point out.

The CMU Center for Machine Translation is also reported to be organisers of a joint international project involving IBM, Hewlett-Packard, and a number of Japanese computer companies (Sigurdson and Greatrex 1987); Carbonell (at MT Summit, September 1987) spoke of the Center's ambition to become a major research centre for MT on a broad front, utilizing AI and knowledge engineering techniques, text understanding, and speech input. It would be building on the existing achievements of Carbonell and his colleagues in AI, computational linguistics and experimental MT. The goals are high quality systems that work, which for MT implies the interlingua approach, and the utilization of specific domain knowledge bases.

Research continues also at New Mexico State University by X. Huang (previously at University of Essex) on an experimental English-Chinese system XTRA using the Definite Clause Grammar formalism (Huang & Guthrie 1986). The project has developed a parsing system which comprises component 'parsers' to test for sentence structure, to try out all possible adjective and noun relations within noun phrases, to check for subject-verb and verb-object compatibilities. The components operate in parallel and can be called interactively by each other. There are clear implications for efficient MT analysis programs.

11.

Experimental MT systems such as these are often short lived. There do not, for example, appear to have been further developments of the AI (or knowledge-based) systems at Yale and at Georgia Institute of Technology - descriptions are to be found in Nirenburg (1987) and summaries in Hutchins (1986: 279-280,282). On the other hand, new AI-inspired systems have been reported; some are no more than toys (e.g. Lee (1987) on English-Korean), others are more substantial, even if still on very small scales.

At the University of California (Irvine), Yoshii (1986) has been experimenting with a Japanese-English system JETR - a small-scale knowledge-based MT experiment, based on a corpus of Japanese cooking recipes and instructions for digital watches. The structural analysis of Japanese is based primarily on the identification of particles, which are treated as indicators of case roles. It combines a top-down processing looking for the fulfilment of predictions and a bottom-up processing to narrow down multiple parses and to resolve slot mismatchings. A major feature is the ability to deal with incomplete and elliptical input, e.g. missing subjects and verbs, missing particles and unknown words, by reference to domain knowledge and an 'inferencer'. JETR distinguishes between information inferred from (incomplete or elliptical) text and information explicitly given in the text. Other knowledge-based MT systems, which do not make this distinction, generate from semantic representations and produce paraphrases rather than translations. Generation of English is in fact performed as soon as there is sufficient information from the Japanese analysis to produce a coherent phrase. Thus, for example, a Japanese phrase translated so far as *'Turn the crown quickly clockwise...'* is changed, when encountering the Japanese subordinator 'to', to *'When you turn the crown quickly...'* In this way, the phrase order of the original is largely maintained, and, it is claimed, the system preserves "the syntactic and semantic content of both grammatical and ungrammatical sentences." Although clearly a knowledge-based approach in that inferencing and role-filling are dominant features, JETR is similar in certain respects to earlier 'direct translation' systems, being designed specifically for two particular languages, intertwining analysis and generation and preserving SL phrase order as far as possible in TL output.

From the University of British Columbia, Sharp (1986) has described a small-scale experiment applying the Government and Binding theory of syntactic analysis. The basic argument is that GB-theory provides a principled foundation for designing grammars which comprise a set of universal features and a set of language-specific features. In a translation system for English and Spanish there are three grammatical components: a universal grammar component (a phrase structure grammar based on X-bar syntax and including rules of Move Affix and Move Alpha) and two language-specific components for English and for Spanish. The latter consist of a lexicon (lexical entries and tables of inflections) and grammar rules characteristic of the language in question; in the case of English it would include rules of Subject-Aux inversion, have-be raising, it-insertion, etc; for Spanish it would include rules for Verb preposing Null subject, etc.

The University of Texas has long been a centre for MT activity in the United States (its major project METAL has been sponsored since 1978 by the Siemens company and is shortly to be available commercially, cf.5 above) The tradition of experimental MT research has been continued with an investigation by Alam (1986) of the potential of Lexical-Functional Grammar as a formalism for MT, with particular reference to problematic aspects of Japanese syntax, and with a study by Jin & Simmons (1986) of the feasibility of writing a single set of bi-directional procedural grammar rules which could accomplish both parsing into a logical form and generation from logical form, and of paraphrase rules to convert from logical form of one language into logical form of another and vice versa. The languages selected were English and Chinese; the goal was thus one single grammar which both parses and generates Chinese and one single grammar which both parses and generates English. Analysis was in the form of deep case 'semantic representations'; paraphrase (transfer) rules mapped SR structures of one language to SR structures of other language. The corpus was very small but it was concluded that "it is definitely possible to write grammars to translate between two subsets of natural languages using symmetric rule forms". (The research has clear affinities and relevance to the Rosetta project, see 17 below.)

A similar concentration on basic MT design is evident in the recent work of two ex-members of the TAUM project (a final summary and survey of the substantial achievements of the TAUM-AVIATION system has been given by Isabelle and Bourbeau 1985/1988). Isabelle and Macklovitch (1986) have begun an experimental English-French and French-English system, written in Prolog, which applies a strictly modular transfer approach. SL analysis is strictly TL-independent, and transfer is limited to strict lexical equivalence with no details of constructional features: thus in a transfer dictionary neither English *know* nor French *savoir* and *connaître* would indicate any structure constraints. (In basic philosophy the approach pursues the rigorous minimisation of transfer seen in the Eurotra specifications, cf.12 below.)

Finally, mention should be made of the revival of MT research at Georgetown University, under Michael Zarechnak (LT4 Dec 87), and the alleged continued interest in MT development by the 'father of Systran', Peter Toma, who is reported to be working on a new system Textus, for English-French and French-English translation.

12.

The EUROTRA project is the largest MT project in the world (in number of personnel, and possibly in expenditure also) and it remains one of the most ambitious and most experimental, in that it is attempting to define the foundations of multilingual high-quality translation.

Eurotra began in 1978, with two basic aims: to construct a prototype MT system for the official languages of the European Community, and to develop expertise in MT and related areas within the Community. The languages are Danish, Dutch, English, French, German, Greek and Italian. The implications of the addition of Spanish and Portuguese are still not clear (Arnold & Des Tombe 1987). The domain has been limited to information technology and official community documents, and the initial prototype is to have relatively small dictionaries of 20,000 items for each language. The project has been extremely ambitious in terms of the number of languages involved, the high quality of system design, and the political logistics.

Eurotra has an estimated 100 full or part-time researchers in 16 different locations: Belgium (Louvain/Leuven, Liege), Denmark (Copenhagen), France (Centre d'Etudes Linguistiques pour la Traduction Automatique, Nancy), Germany (Saarbrücken, Stuttgart, Berlin, Bielefeld) Netherlands (Institute for Applied Linguistics, Utrecht), Greece (Athens), Ireland (Dublin Institute of Technology), Italy (Institute of Computational Linguistics at the University of Pisa), Luxembourg, United Kingdom (University of Essex, University of Manchester Institute of Science and Technology). In addition there is a Central Unit of about 12 specialist linguists, computer scientists and translators, and a secretariat provided at ISSCO (Geneva). The project is administratively complex thanks to the deliberate policy of decentralisation.

A general account of the current status of the project is provided by Arnold (1986), Arnold et al. (1986), Arnold & Des Tombe (1987) and by a special issue of Multilingua (Somers 1986b), which includes accounts of each of the language teams. Earlier stages are represented by Johnson et al. (1984, 1985) and are summarised in Hutchins (1986). Other speculative theoretical contributions of a more specific nature are given by Schmidt (1986) on valency, by Hauenschild (1986) on AI approaches, and by Rohrer (1986a, 1986b) on unification grammar.

The Eurotra project has always stressed the need for explicit and appropriate formalism (e.g. Johnson et al. 1984), firstly for the practical reason that decentralisation of research requires each team to be fully clear of the framework in which it has to work, and secondly for the theoretical reason that multilinguality demands a level of abstraction not previously attempted by any MT project. In the course of the last few years the previous dependence on a GETA-type formalism has been dropped, and there has now emerged a more firmly based theory which embodies the strict application of the compositionality principle, a unification grammar, lexical-functional grammar ideas, etc. The new <c,a>t framework is intended to overcome problems of the older unconstrained 'standard', which allowed too much latitude to alternative formulations (Arnold et al. 1986). The framework is designed to conform to the basic theoretical principles of differentiation (i.e. that meaning distinctions are preserved), simplicity, specificity (inc. perspicuity to researchers) and 'isoduidy' (i.e. that structures must have the same interpretation).

The transfer approach is seen as not only the most practically feasible design for MT, but also as the most appropriate from theoretical considerations. The basic premiss is that the translation process should be regarded as a series of relations between representations (of sentences and texts) which are necessarily linguistic in nature. Translation involves more than the preservation of meaning (unlike 'paraphrase'); in MT representations must be linguistic, and less abstract than content representations. It is argued that the representation languages at the interfaces of analysis and transfer and at the interfaces of transfer and generation have to be formalisms which, having resolved language-specific 'ambiguities', are capable of distinguishing all those different interpretations leading to different translations; furthermore, the representation languages must be readily understood and easily learnable by linguists working on the system (i.e. they must be theoretically 'coherent' formalisms). The latter requirement is achieved by defining representation languages as generative devices ('grammars'), which evaluate for 'well-formedness'. A particularly pressing requirement in a multilingual system is that the transfer component should be as 'simple' as possible; consequently, further levels of representation are necessary within SL analysis and TL generation (giving a 'stratificational' model) - at present five levels are being suggested for each language: case frame structure (the interface semantic dependency structure), relational syntax (equivalent to LFG f-structures, and including surface grammatical relations such as subject, object, etc.), configurational syntactic structure (surface structures as represented e.g. in GPSG), morphological structure (decompositions of word structures), and normalised text. Relations between representations are defined by 'translators', a set of rules which are constrained by the principles of 'one-shot'-ness and compositionality. The first principle requires simply that no intermediary representations are created. The second states that the interpretations of structures are functions of the interpretations of their components in a formally defined way. By the notion of compositionality is meant that the translation of a complex expression should be a function of the translation of the basic expressions it contains together with their mode of combination - in this respect Eurotra shows the influence of Landsbergen's Rosetta project (below), whose influence on the Eurotra formalism is readily acknowledged. In Eurotra, however, there is uncertainty whether 'translators' should operate between representations or (as in Rosetta) between derivation trees. The

Eurotra approach is also less stringent with respect to the 'isomorphism' of grammars, i.e. it does not extend to the requirement that grammars should be 'attuned' to each other and be truly isomorphic.

The main grammar formalism is the nondeterministic tree-tree transducer (as found also in TAUM and GETA); representations are in the form of dependency trees, at all levels: morphological, surface syntactic, deep syntactic, and interface. The latter are, as in GETA, multi-level representations including morpho-syntactic features as well as semantic features. In general, the formalism is regarded as equivalent in power to that of lexical-functional grammar (Arnold 1986)

There can be no doubt that the theoretical activity of Eurotra researchers has contributed substantially to the theoretical foundations of MT. However, it is conceded readily and frankly that "the results of this research are not very clear yet" and that the obvious limitation is "the lack of a principled account of the lexicon" (Arnold & Des Tombe 1987). Eurotra remains an explicitly 'linguistic' system; it does not attempt to incorporate any knowledge bases or cognitive modelling of the AI kind. It has, therefore, been obviously open to strong criticism for its concentration on abstract formalism, for neglecting the construction of actual grammars, parsers, generators and dictionaries, its insufficient empirical testing, and what is seen as a 'narrow' exclusion of discourse phenomena and advances in knowledge based approaches to natural language processing. Its design was beginning to look 'obsolete', a batch system with no interactive component, an exclusively 'linguistic' model with no AI. In addition, the project was behind schedule; a small prototype was very slow (over 20 minutes for one sentence!); and it had received about $40million from the CEC and contributing countries (LT1 June 1987). The European Parliament set up an independent evaluation.

The report appeared in late 1987 (the Pannenborg report, CEC 1987). It was certainly not uncritical: problems of management had not resolved; there was a lack of central resources; the exclusive emphasis on linguistic foundations had led to a neglect of computational possibilities (e.g. of interactive approaches); there had been insufficient attention to dictionary compilation. The project was criticised for excessive research and for straying from its mandate to create an operational system; the report confirmed fears that the project will fail to produce a prototype by 1990 or a commercial system by 1993. Nevertheless, the project had made fundamental progress in the specification of interfaces and had succeeded in promoting computational linguistics research in member countries. It concluded that it would be a retrograde step for CEC to abandon the project at this stage; what was needed was more realistic deadlines, and it recommended that the research phase should come to an end in 1988 and that the project should then establish closer links with industry and start development of a practical system.

13.

The French MT research group at Grenoble is now the longest established. In the 1950's and 1960's its CETA system explored the interlingua approach to MT system design; since 1971 the GETA project has done some of the most fundamental research on the transfer design. The latest version Ariane-85 is a further development of the now well known Ariane-78 system described, e.g. by Boitet (1984/1987), Vauquois & Boitet (1985/1988) and summarised in Hutchins (1986: 239-247). The fullest descriptions of Ariane-85 are given in Boitet 1986, 1987a, 1987b. The basic design retains the modular structure, separate SL, TL and bilingual dictionaries and grammars, in the following stages: morphological analysis (strings to trees), structural analysis (producing abstract multilevel tree representations which combine syntactic, logical and semantic relationships), lexical transfer (lexical substitution with some structural changes), structural transfer (tree transduction), syntactic generation, morphological generation (trees to strings).

The long-term aim of the GETA project is a multilingual system producing 'good enough' results, i.e. accepting the need for post-editing. The system is essentially, like Eurotra, a linguistics-oriented system; it does not claim to use any 'deep understanding' or 'intelligence', and hence no AI-type explicit 'expertise' is incorporated in GETA-ARIANE - although the possibility of grafting on an 'expert' error correction mechanism was investigated by Boitet and Gerber (1986). However, unlike other linguistics-based systems, Ariane extends translation analysis to sequences of several sentences or paragraphs, in order to deal with problems of anaphora and tense/aspect agreement. For practical production the system permits optional pre-editing, primarily the marking lexical ambiguities; post-editing can be done using the REVISION program developed for ARIANE-78. It is a mainframe batch system with no human interaction during processing. However, Zajac (1986) has investigated an interactive analysis module for GETA, somewhat on the lines of Tomita's research at Carnegie-Mellon (Tomita 1986).

One important development has been the refinement of the theoretical basis, particularly the clarification of the distinction and the relationship between dynamic and static grammars in the system. Static grammars (or SCSG 'structural correspondence static grammars') record the correspondences between NL strings and their equivalent interface structures in a formalism which is neutral with respect to analysis and synthesis. The processes of analysis and generation are handled by 'dynamic grammars' written in appropriate 'special languages' (SLLPs or Special Languages for Linguistic Programming): ATEF for morphological analysis, ROBRA for structural analysis, structural transfer and syntactic generation, EXPANS for lexical transfer, and SYGMOR for morphological generation. (The distinction between 'static' and 'dynamic' grammars is now found in many advanced transfer systems; the GETA project has been a leading force in this theoretical development.)

Equally important have been the improvements to the research environment, in tools for the development of systems, such as ATLAS for lexicographic work and VISULEX for viewing complex dictionary entries. Such tools are components of a 'linguistic workstation' for MT research (an idea also being developed by the Saarbrücken and the Kyoto groups, 15.and below). Within this environment the work of the Calliope project has taken place: the compilation of the static grammars for English and French during 1983-84, their corresponding dynamic grammars, and the substantial lexicographic work.

The Grenoble group has always encouraged and supported other MT projects using GETA software, and thereby helped to train MT researchers. ARIANE is regarded above all as "an integrated programming environment" for the development and building of "a variety of linguistic models, in order to test the general multilingual design and the various facilities for lingware preparation..." The ARIANE software has been tested on an impressive range of languages, often in small-scale experiments (Vauquois and Boitet 1985/1988; Hutchins 1986: 247-8; Boitet 1987a), but sometimes in larger projects, e.g. the English-Malay project mentioned elsewhere in this survey.

The largest GETA-ARIANE system has been for Russian-French translation, which built upon previous experience with CETA. Since 1983 this system has been extensively and regularly tested in an experimental 'translation unit'; large corpora of text have been translated, including some 200,000 running words during one 18-month period (Boitet 1987b). Another large-scale system was the German-French system developed by Guilbaud and Stahl, using the same generator programs as in the Russian-French system. Its principal features were the attention given to morphological derivation and inflection, and the restriction of structural analysis almost wholly to morphological and syntactic data, with little or no use of semantic information. The system has been described by Guilbaud (1984/1987), but there has been little development of the system since 1984 (Boitet 1987b).

The most important practical application of a large-scale system has, however, been through GETA's involvement in the French national computer-assisted translation project (NCATP). Launched in November 1983 (after a preparatory stage in 1982-83, the ESOPE project), the Calliope project has been financed 50% from public funds (administered by the Agence d'Informatique) and 50% from private sources. One source has been B'VITAL, founded in 1984 by the Grenoble group, which is responsible for the machine-readable dictionaries and for the 'static grammars' (Joscelyne 1987). Another has been Sonovision, which was to provide the aeronautics terminology for the major French-English system Calliope-Aero. After a demonstration of a prototype of Calliope-Aero at Expolangues in February 1986, it was decided to develop also an English-French system for the translation of computer science and data processing materials, Calliope-Info. In addition to these MT systems, both batch systems, the project was also to produce a translator's workstation (Calliope-Revision, organised around a Bull Questar 400 microcomputer) for preparing and post-editing texts and for access to remote term banks and including OCR and desk-top publishing facilities. This was essential if the systems were to be fully integrated into an industrial documentation environment. However, given the expected delays there have been plans by SG2 (one of the backers) to develop a terminology aid with split-screen word processing, Calliope-Manuel.

Whatever the commercial feasibility of the Calliope project, which came to a formal end in February 1987 (Boitet 1987a), the experience will no doubt be put to good use by the GETA project, in particular the experience of dealing with complex dictionaries and the type of scientific and technical sublanguage presented by aeronautics. Boitet (1986), for example, mentions the successful treatment of complex noun phrases (e.g. *la jonction bloc frein et raccord de tuyauterie*) and complex adjectival phrases (e.g. *comprise entre les deux index noir*). Other problems did not occur in the sublanguage and thus were put aside, e.g. interrogatives, relative clauses introduced by *dont*, imperatives, certain comparatives, nominal groups which do not only consist of nouns, and so forth.

The NCATP has had other consequences. It stimulated the conversion of the ARIANE-85 to run on IBM PC AT (with a minimum 20MB hard disk), adequate for MT development but not for a production system. It also encouraged the writing of new software in a French dialect of LISP (Boitet 1986; Boitet 1987a, 1987b), with the aim of creating a fully multilingual system with a single 'special language' for processing strings and trees (TETHYS). Clearly, GETA has continued to advance the boundaries of MT research.

14.

While GETA is the main MT research centre in France, there are other MT projects in Nancy and Poitiers. At Nancy, Chauché (1986; Rolf & Chauché 1986) continues his research, begun at Grenoble, on algorithms for tree manipulation which are suitable for MT systems. Tests of the algorithms have been applied to Spanish-French and Dutch-French experiments (in collaboration with Rolf of Nijmegen University). From Poitiers, Poesco (1986) reports a small-scale knowledge-based MT experiment for translating Rumanian texts on three dimensional geometry into French. The ATN parser produces a conceptual frame-slot representation from which the generator devises a 'plan' for producing TL output.

The restricted language system TITUS, designed for multilingual treatment of abstracts in the textile industry, has expanded in its latest version TITUS IV (Ducrot 1985) in order to deal with a wider range of subjects and to allow somewhat freer expression of contents.

As elsewhere, there is commercial interest in translators’ workstations: Cap Sogeti Innovations is proposing a "language engineering workshop", providing 'intelligent' language tools, a dedicated multilingual word processor, a natural language knowledge base, a technical summary writer, and a 'text analyzer' which will produce abstract meaning representations. Details are necessarily vague at present (Joscelyne 1987).

Attitudes to MT in France are most likely to be changed by the provision of MT services on Minitel. The availability of Systran has already been mentioned (sect.1 above). Other services include a number of dictionaries and term banks: the Harrap French and English slang dictionary, the Dictionary of Industries, Normaterm (the term bank of the French standards organisation

AFNOR), the DAICADIF lexicon for telecommunications, and (next year) FRANTEXT the historical dictionary Tresor de la Langue Française.

15.

The largest and most long-established MT group in Germany is based at Saarbrücken. It began in the mid-1960's with research on Russian-German translation, sponsored from 1972 to 1986 by the Deutsche Forschungsgemeinschaft. The SUSY project expanded into a multilingual system, based on the transfer approach, with the source languages German, Russian, English, French, and Esperanto, and the target languages German, English and French. Detailed descriptions of the latest version SUSY II as at the end of 1984 are given by Maas (1984/1987) and by Blatt et al. (1985), and summarised by Hutchins (1986: 233-239).

The most recent developments of MT research at Saarbrücken are to be found in Zimmermann et al. (1987). The most significant are the changes introduced into the basic design by the introduction of English as a SL (in SUSY-E project), the development of explicit formalisms and software tools for testing natural language processing and MT models and for general computational linguistic experimentation (SAFRAN: Software and Formalism for the Representation of Natural Language - see Licher et al. 1987), the planned application of SUSY as the foundation of a production-oriented system (STS), the new direction of Saarbrücken MT research in the ASCOF project for French and German translation, and the involvement of SUSY personnel in the Eurotra project.

The greater emphasis on product-oriented research has arisen in part from the ending of direct DFG funding in 1986. The project MARIS (Multilinguale Anwendung von Referenz-Informationssystemen) was established in mid-1985 at the University of the Saar to develop a multilingual information retrieval system, in particular to meet the needs of German-speaking users of English-language documentation (Zimmermann et al. 1987; Luckhardt 1987b). For this purpose, the MARIS team is developing a computer-assisted translation system STS (Saarbrücker Translationsservice) based on the Saarbrücken MT research. Initially only English-German will be developed, and it will restricted to the translation of abstracts and titles of journal articles. There are three phases to the project: first a manual system in which translators can have access to computer-based term banks, secondly the addition of automatic lookup of terminology, and thirdly the application of SUSY as a post-edited MT system. The chief emphasis will be on lexicographic data and sublanguage information in the particular fields of application: housing and construction, environment, standards, social sciences. At a later stage it is hoped to add French-German and German-French versions. The MARIS project is a natural continuation of basic MT research on SUSY, some earlier experience with a prototype translator's workstation (SUSANNAH) and the long-established research at Saarbrücken under Harald Zimmermann on information retrieval systems.

The ASCOF project (Projektbereich C at Saarbrücken) grew out of research at Saarbrücken on computational methods for the analysis of the 'Archive du Français Contemporain' (established in mid-1960's). The project team initially worked in close collaboration with GETA. However, from the mid 1970's, after the elaboration of French analysis programs for SUSY, the team established closer links with the Saarbrücken MT projects. For a while the research was using both GETA and SUSY algorithms, but since 1977 it has concentrated on SUSY-type methods, and in 1981 it emerged as an independent MT project at Saarbrücken (Scheel 1987). Its distinctive features are: the use of the COMSKEE programming language, the integration of syntactic and semantic analysis, the adoption of ATN parsers, and the use of semantic networks as a 'knowledge base' for disambiguation.

ASCOF (Biewer et al. 1985/1988; Stegentritt 1987) is a system for French-German translation with a multilevel modular transfer design. Programs are written in COMSKEE (Computing and String Keeping Language), the programming environment developed at

Saarbrücken. The object of ASCOF (=Analyse und Synthese des Französischen mit Comskee) is fundamental MT research not a practical system.

Analysis is in three basic phases. The first phase of morphological analysis is followed by a second phase in three parts: disambiguation of word class homographs, identification of non-complex syntactic groups, and segmentation of sentences into independent (unrelated) parts, e.g. noun, verb and prepositional phrases. In the third phase, structural analysis is realised by a series of cascaded ATN parsers which combine syntax (e.g. functional relations) and semantics (e.g. case frames or valencies), with no priority given to either one or the other. Analysis modules operate not sequentially but interactively: thus analyses of verb phrases, complex noun phrases, complements, coordination etc. interact with each other. The integration of syntax and semantics in ASCOF analysis contrasts with procedures in SUSY and other linguistics-based transfer systems. Lexical disambiguation is achieved by reference to semantic networks which include information on synonymy, homonymy, hyponymy (e.g. whole-part, genus-species), and semantic-functional frames for verbs. As in GETA and EUROTRA, the results of analysis are not interlingual representations but SL canonical trees in which SL-specific lexical and syntactic ambiguities have been resolved.

Transfer operates in a familiar way with bilingual lexical substitution and structural tree transduction, and it is followed by TL syntactic synthesis and morphological synthesis for the production of TL text output. Most of the ASCOF research effort has concentrated on problems of analysis and on testing the semantic network approach to disambiguation. Consequently the transfer and synthesis components have not yet been fully developed; there is no French synthesis program and only a small German one, so only partial implementation of translation from French into German has been possible so far (Stegentritt 1987). The system is to be tested on EEC agricultural texts, and this corpus has provided the data for the illustrative semantic networks. The quality of the output is considered to be crucially dependent on the development and elaboration of the semantic networks, the linguistic 'knowledge base' of the system. ASCOF is an example of a transfer system of the third generation of MT, incorporating AI-style 'knowledge base' semantic analysis, and aiming in the long-term for high quality (batch) translation.

16.

ASCOF has not been the only context in which research on knowledge based approaches to linguistic analysis has been conducted at Saarbrücken. There has been activity in text-oriented MT by Weber and Rothkegel. Weber (1986, 1987) has investigated a small-scale text-oriented system for MT. COAT (=Coherence Analysis of Texts) is a program, written in COMSKEE, which establishes text coherence, using information about valencies, arguments and roles, and produces complex representations of SL texts which might be translated into equivalent TL text representations. Analysis is not to the depth of AI-type understanding, only sufficient for translation. A speculative extension is OVERCOAT (not implemented) for establishing global text structures (paragraph sequencing) and AI-type discourse frames, and involving 'knowledge' of stereotypical situations and events. Rothkegel's (1986a, 1986b, 1987) research between 1981 and 1986 was devoted to TEXAN, a system for recognising text-linguistic features (illocution, thematization, coherence, etc.), for identifying text types (specifically EEC treaties) and consequently enabling text-specific semantic and structural analysis, disambiguation and transfer.

The Saarbrücken group had collaborated with Kyoto to develop a system for translating German journal titles into Japanese. SUSY was used for the analysis of German titles and TITRAN for the generation of Japanese titles (Ammon & Wessoly 1984-85). There has subsequently been a similar project at Stuttgart to produce German translations of Japanese journal titles in the information technology field (Laubsch et al. 1984, Rosner 1986a, 1986b). The SEMSYN (Semantische Synthese) system takes as input the semantic interface representations of Japanese texts produced by Fujitsu's ATLAS/II system (cf. 30 below). ATLAS was designed for Japanese-

English translation and so the semantic interface representations were not completely sufficient for German synthesis, since they gave few indications of number, definiteness, or tense. SEMSYN is a semantics-based MT system (or rather partial system); it incorporates a frame description formalism (cases, roles, modalities, scopes, purpose, part-whole relations, etc.) from which German titles are generated by reference to a restricted knowledge base of linguistic and extra-linguistic information. Like many AI-inspired systems, SEMSYN is written in LISP.

The Saarbrücken group was an early participant in the Eurotra project and it has contributed a number of theoretical studies. Two recent examples are the work of Steiner (1986) on generation and of Schmidt (1986) on valency structures. However, the West German Ministry for Research and Technology (Bundesministerium für Forschung und Technologie) has also established three groups in Berlin, Bielefeld and Stuttgart to undertake theoretical research on behalf of the Eurotra project. The BMFT project as a whole is known as NASEV (Neue Analyse- und Syntheseverfahren zur maschinellen Übersetzung).

At Stuttgart, Rohrer (1986a, 1986b) has been investigating the relevance of formal linguistics to the theoretical basis of MT. He advocates unification grammars (e.g. LFG, GPSG, FUG) as offering the most appropriate general frameworks for future advances in MT research. At the Technical University of Berlin, Hauenschild (1986) has been investigating AI approaches to problems of MT transfer. This research, under the acronym KIT (Künstliche Intelligenz und Textverstehen), commenced in April 1985, and can be regarded as a continuation of her previous work on the CON3TRA project at the University of Konstanz and on the earlier SALAT project at Heidelberg. Hauenschild's model proposes (i) SL analysis in terms of a modified Generalized Phrase Structure Grammar, (ii) conversion into an intentional logic representation directly from the GPSG analysis (applying compositional semantic rules in the manner of Montague grammar), and then (iii) conversion into two levels of semantic representation: a level of 'referential nets' linking text referents, and a level of global 'text argument' structures recording intersentential relations. Transfer would operate at multiple levels: lexical (at semantic representations), sentence-semantic (at intentional logic representation, i.e. in order to preserve informational structure of SL texts), and syntactic (i.e. from 'superficial syntactic' GPSG analyses). The semantic representation language is a propositional logical formalism (with variables and operators), and includes knowledge of facts, rules and objects. The 'argument structure' representation is regarded as genuinely interlingual in so far as logical, case and argument features may be 'universal'. However, the precise division of levels is still fluid. As a MT model, Hauenschild's work represents the convergence of many recent strands of MT theory.

17.

Two of the most innovative MT projects at present time are based in the Netherlands. Both have chosen to develop interlingua models. At Philips in Eindhoven, the Rosetta project is exploring a system based on Montague grammar; at the BSO software company in Utrecht, the DLT project is building a system with Esperanto as the interlingua.

The basic principles of Rosetta derive from its foundations on Montague grammar (Landsbergen 1984/1987, Appelo & Landsbergen 1986, Leermakers & Rous 1986, Landsbergen 1987). The main characteristic of Montague grammar is the derivation of meaning representations (interpretations) from the syntactic structure of expressions. A fundamental principle is compositionality, namely the premiss that the meaning of an expression is a function of the meaning of its parts. Since the parts are defined by syntax, there is a close relation between syntax and semantics. The link is the correspondence of syntactic derivation trees and semantic derivation trees. Syntactic derivation trees represent the processes by which syntactic rules are applied to produce a syntactic analysis (parsing) of a sentence. For each rule of a syntactic derivation tree there is taken to be a corresponding semantic operation; hence, the semantic value of a full syntactic derivation is given by its corresponding parallel semantic derivation tree. From such semantic derivation trees may be derived logical expressions (in an intentional logic formalism), and this is frequently the preferred option by Montague grammarians.

One model for a MT system based on Montague grammar would, therefore, be to use these logical expressions as interlingua representations. But this would entail the loss of information about the surface 'form' of messages or texts, and this is information which can be vital for generating satisfactory translations. Furthermore there would be the problem of devising a single logical formalism for a wide variety of languages. Consequently, the Rosetta project has taken a different approach: it aims to use the semantic derivation trees as interlingual representations. This is done by making the syntactic derivation trees of the languages in the system isomorphic, and isomorphism is achieved by attuning the grammars of the languages in question (their 'M-grammars') so that for every syntactic rule in one language there is a corresponding syntactic rule in the other with the same meaning operation, i.e. so that the processes of constructing or deriving sentences in one language are parallel to the processes of constructing and deriving (translationally) equivalent sentences in the other language. Thus the corresponding semantic derivation trees are identical and in effect interlingual representations for the languages whose grammars have been 'attuned' appropriately. Rosetta is thus intended as a multilingual interlingua system; initial research (in the six-year project starting in 1985) will concentrate on developing successively more sophisticated M-grammars of three languages, English, Dutch and Spanish; with other languages to be added at later stages.

Rosetta adds another principle to those of compositionality and isomorphism. This is the reversibility principle: a single grammar of a language (M-grammar) should be the basis of procedures for both analysis (M-parser) and synthesis (M-generator), so that for each analytical rule there is a reverse generative rule, i.e. grammar rules should be reversible in a bi-directional MT system. Explicitness and rigour of grammars, formalisms and theoretical foundations are natural and expected concomitants of a MT model based so firmly on achievements in formal linguistic theory. Rosetta is, then, deliberately and explicitly a linguistics-oriented model. Specific linguistic problems and their treatment have been reported: temporal verbs (Appelo 1986), idioms (Schenk 1986), and synonymy (de Jong & Appelo 1987). The need for extra-linguistic knowledge is recognised in a practical translation system, but the incorporation of AI-type extra-linguistic data in the model is to follow at a later stage of the project.

18.

The MT project in Utrecht at the software company, Buro voor Systeemontwikkeling (BSO) began in 1982 with a feasibility study supported by the EEC. BSO set up a six-year project in 1985, with the assistance of a substantial grant from the Netherlands Ministry of Economic Affairs, to build a prototype system for translating from English into French, with a commercial version expected in 1993.

DLT (Distributed Language Translation) is designed as an interactive multilingual system operating over computer networks, where each terminal acts as a translating machine from and into one language only; texts are transmitted between terminals of the network in an intermediary language, a version of Esperanto. DLT is not, therefore, a 'translating machine' or a tool for translators but primarily a tool for interlingual communication, enabling monolingual users (authors) the means to generate their texts in other languages. Users will know only the language of source texts; hence they will interact during analysis and transfer in their own language - in some cases, interaction may lead to rephrasing original texts in order to remove ambiguities of translation problems. A description of the present DLT system is given by Schubert (1986), and more extensive treatments of the semantics and syntax in Papegaaij (1986) and Schubert (1987).

As an interlingua system, DLT has two basic parts: analysis of SL texts into IL (Esperanto) representations, and generation of TL texts from IL representations. Analysis is by far the most complex, since the results must be unambiguous for both the SL and for any of the potential TLs. In the DLT system the decision has been made to restrict linguistic analysis of SL sentences to morphological and syntactic features and to concentrate all semantic disambiguation to IL stages. There are therefore no SL and TL semantic components, all semantico-lexical knowledge is represented in an IL (Esperanto) database. The arguments in favour of Esperanto as an IL are given as being its NL-like expressiveness and flexibility, its regularity and consistency, and its independence from other NLs (its autonomy) - i.e. unlike the IL in Rosetta, the DLT IL is not 'attuned' (adapted or refined) to the SL and TL involved.

Analysis passes through the following stages: SL syntax parsing, SL-IL transfer, IL semantic 'word expert' system (SWESIL), SL dialogue, IL linearization. In the DLT English-French prototype texts are composed in Simplified English (a 'restricted grammar'), with the aim of free unrestricted input by 1990. Syntactic analysis is performed by an ATN parser producing dependency tree representations; multiple analyses are common because no resolution of non-syntactic homography or ambiguity is undertaken. SL (English)-IL transfer involves three aspects: (a) replacement of English words by tentative IL words, (b) arrangement of IL lexical items in syntactically correct IL trees, and (c) selecting the semantically and pragmatically best tree. The first step is done by straightforward substitution via a bilingual English-IL dictionary. The second step is done by 'metataxis' rules which take as input subtrees relating groups of (English) lexical items and produce as output their corresponding IL subtrees; a detailed account of metataxis and its linguistic foundations is given by Schubert (1987).

At this stage, there will be a number of IL subtrees representing different interpretations of the SL (English) input sentence. It is the task of the IL 'word expert' system to determine which is the correct or most probable interpretation. SWESIL (Semantic Word Expert System for the Intermediate Language) tests pairs of IL words linked by a dependency relationship (expressed by an IL relator, e.g. preposition) for meaning compatibility, computes a score which indicates the probability of the given two words occurring in the given dependency relation, ranks the probability scores for all the possible relationships and determines which meanings and which relationships are to be selected. In this way SWESIL resolves syntactic ambiguities of SL input which cannot be solved by monolingual syntactic information, and at the same time it defines unambiguous IL interfaces from which TL forms are generated. The knowledge required by SWESIL is both syntagmatic, i.e. lists of acceptable IL word pairs and their relators, and paradigmatic, i.e. taxonomies of IL words, tree-structures of part-whole, genus-species, hypernym-hyponym relations, etc. It also contains both linguistic and world knowledge, and is thus modelled on AI-type knowledge-based systems, although unlike many systems DLT does not use semantic features' or 'primitives'. (More details of SWESIL are given by Papegaaij 1986.)

Any residual ambiguities are presented for resolution to the user in a computer-initiated interactive dialogue in the user's own language, i.e. not in Esperanto. (It is intended that the results of these interactions will be incorporated in the SWESIL knowledge base, which will thus become a 'learning' system.) Finally, the 'unambiguous' IL text is linearized as quasi-Esperanto sentences for transmission in the network. At earlier stages of the DLT project it was anticipated that Esperanto would have to be extensively modified; in practice little modification has been found to be necessary and so the linearized IL representations are consequently readily understood by anyone familiar with 'normal' Esperanto (thus facilitating inspection of DLT system performance by researchers).

The conversion of the IL text into TL text is to be performed without any human intervention. Synthesis is the inverse of analysis: problems of TL word choice are resolved by reference to SWESIL information, a bilingual IL-TL (French) dictionary and a set of IL-TL metataxis rules converts IL structures into TL (French) dependency trees, which is in turn linearized as French sentences.

A launch of the DLT prototype in December 1987 (LT5; LM52), though based on a small 2000 word vocabulary, demonstrated the basic feasibility of the approach; the present operational slowness is expected to be overcome by future advances in parallel processing. Most research effort will concentrate on building the Esperanto based knowledge bank. The DLT group has recognised the importance of tackling lexical problems of MT from the beginning, instead of leaving them (as many MT projects have done) until after procedures for morphological and syntactic analysis and synthesis have been fixed. For this reason, an initial test of the lexical adequacy of the system was undertaken independently by Melby (1986; Sadler & Papegaaij 1987).

19.

Although there is no large MT project in Belgium there is an impressive amount of activity in this area and in the related fields of computational linguistics and artificial intelligence. Researchers at the universities of Liège and Leuven (Louvain) in Belgium participate in both the Eurotra project (the Belgo-Français and the Belgo-Dutch groups respectively) and in the development of the Siemens METAL system for Dutch-French translation (which also has input from the University of Mons). In addition, there is an independent project at the Free University of Brussels to develop a microcomputer system for English-French translation of computer manuals (Luctkens & Fremont 1986). The ultimate aim is a multilingual transfer design, but as yet the experiment has concentrated primarily on developing an ATN parser for syntactic analysis.

20.

In the recent past MT research in Great Britain was negligible; now there are some signs of greater interest and support. The centres for the Eurotra project are the University of Essex and the University of Manchester Institute of Science and Technology (UMIST). Both centres have made substantial and important contributions in the areas of software development, linguistic theory, and the formalization of the basic environment. The MT experience at UMIST has been augmented by a grant from the governmental Science and Engineering Research Council (under the so-called Alvey initiative), and with substantial support from the computer company ICL. A joint project has been set up with the University of Sheffield: the UMIST team are undertaking research on an English-Japanese system, the Sheffield group on a Japanese-English model. Knowles (1987) gives a general outline of the two projects, due to come to an end in October 1987.

The English-Japanese prototype (NTRAN) at UMIST is an interactive transfer system, written in Prolog with Lexical Functional Grammar providing the basic framework for parsing, transfer and generation. A bottom-up parser produces LFG-type F-structures; from these are derived the S-structure interface representations which are converted into equivalent Japanese interfaces; these S-structures are the source of the F-structures for Japanese surface strings.

In this project the approach to interactive MT differs from the familiar one (although shared by the DLT project). It is argued that if a MT system should concentrate on producing good TL output on the basis of SL texts interpreted interactively with users of the system (Johnson and Whitelock 1987; Whitelock et al. 1986). The user is thus freed from the need to know anything about the TL. Any deficiencies of the MT system in knowledge of the SL or in the subject matter of the texts will be made good by the user in computer-initiated dialogues (on lines indicated by Tomita 1986). The intention is that users of the system would be authors who compose documents in their own language (English) and are prompted for explanation (and resolution) of ambiguities in their own language. Disambiguation would take place not during analysis but during transfer, when it will be known not only what SL ambiguities are present but also what particular translation problems are prompted by the language-specific characteristics of the TL. Although the aim is essentially to solve translation difficulties, it is envisaged that the mechanism could also act as 'intelligent style-checker', i.e. an interactive rewriting component. NTRAN is deliberately underdetermined in SL syntax and lexicon semantics; the emphasis is on the knowledge base of linguistic and extra-linguistic information necessary for transfer and generation. In this way, the 'expertise' of the MT system resides in its 'knowledge' of the TL and the problems of interpretation from the perspective of that language. It is not an 'expert' in the SL and the subject; that knowledge is provided interactively by the operator (or author).

The Sheffield (AIDTRANS) model for Japanese-English translation is derived from research on teaching people with little linguistic talent how to 'decode' Japanese texts by using a sophisticated grammar-dictionary in an 'automaton-like' fashion. (The integrated grammar-dictionary was devised by Jiri Jelinek who had worked on MT in Czechoslovakia before emigrating to Britain.) The system is strictly uni-directional (Japanese-English) and, in terms of MT design, based on the 'direct translation' model of linear predictive analysis. The intention of the project is to computerise the almost 'mechanical' procedures of the human learners ('decoders'), and to produce a type of machine-aided system of translation.

As elsewhere, there are in Britain small-scale experiments. One is the SLUNT system, an interlingua model based on numerical coding of vocabulary items, which is described with a demonstration microcomputer program in Goshawke et al. (1988). Another is TRANPRO, a translator's aid (essentially a multilingual word processing package) providing access to remote dictionaries and in-house glossary updating (LM34 July 1986). A third is the research reported from ICL (International Computers Ltd.) on Telex message translation for the language pairs English and French and English and German (LT5 Jan 1988).

More widely reported is the research, first announced in August 1987, by British Telecom on a speech translation system for business telephone messages (Stentiford & Steers 1987). The system, first announced in August 1987, has been under development since 1984; a marketable system is hoped for in the mid 1990's. The speaker utters his message clearly and deliberately into a microphone attached to a Merlin 2000 microcomputer. The message is confirmed by a synthetic repetition, then translated and spoken by a speech synthesis program in a TL. The languages involved at present are English, French, German, Italian, Spanish and Swedish. The corpus is limited to a set of 400 common business phrases involving some 1000 different words. The computer program is designed to recognise (by pattern matching) 100 keywords from which it can select the intended phrase and generate (synthesise) a spoken phrase in another language. Speech recognition is prone to considerable error, but these can be reduced by restricting the domain and by limitation to a small number of keywords to be identified. The researchers claim that for this limited corpus, any phrase can be identified by three or four keywords (not necessarily 'content words' only). At present, the speech recognizer has to be trained for each individual speaker; the aim is for the system to be speaker-independent, to handle a larger vocabulary and to produce more natural output.

21.

Spain is a newcomer to MT research. Mention has already been made of activity in connection with Siemens' METAL project (sect.5 above). There are recent reports (LT2 August 1987) that the IBM Research Centers in Spain and Israel will be directing a study of the feasibility of MT systems for English-Spanish and English-Hebrew (probably batch systems with post-editing), for use within IBM itself. The company had earlier invested heavily in a trial of ALPS machine aids for English-French translation. The principle aim will be "to develop internal IBM knowledge and expertise in machine translation", involving other IBM Centers in Europe and inviting university participants (Helsinki is already committed). It appears that the system will be written in Lisp and will not be IBM-hardware specific, and that it will probably use the English parser developed at the Yorktown Center in the United States for the EPISTLE system.

22.

The Dalle Molle Institute for Semantic and Cognitive Studies (ISSCO) in Geneva, Switzerland, has been active in the Eurotra project from the beginning, functioning for a while as the secretariat of the project and convening seminars on MT and related aspects of computational linguistics (e.g. the proceedings of the 1984 Lugano tutorial edited by Smith 1987). The Institute has also undertaken experimental work itself, for example, the project described by Buchmann et al. (1984). This is a prototype MT system (SEPPLI) for translating job advertisements from German into French and Italian. The corpus is limited to posts in administration taken from a Swiss government weekly publication. The objective of the project is to define the depth of linguistic analysis necessary in this particular domain and sublanguage - which exhibits a high proportion of noun phrases, few finite verbs and no relative phrases and dependent clauses. The hypothesis is analysis can be relatively 'shallow', with dependency trees representing only partial analyses. For example, there is no attempt to establish the correct relationships of prepositional phrases, coordination, etc. since the surface linearity of the SL texts is retained in the TL output. The limited lexical domain of the sublanguage obviates the need for a semantic component. The model - a familiar transfer design, with independent modules for SL analysis and TL generation - is seen primarily as a didactic and research tool; no large-scale operational system is planned.

23.

Scandinavian activity has been mentioned already: the involvement of researchers at Copenhagen University in the Eurotra project, the ENTRA (English-Norwegian) system developed in collaboration with Weidner (WCC) at Bergen University, the report of collaboration by researchers at Helsinki University in the IBM project (cf.21 above), and the use of MT for translating Japanese databases at the Research Policy Institute of Lund University (Sigurdson & Greatrex). Mention should also be made of the experimental work by Sigurd, also at Lund, on the multilingual SWETRA system.

24.

In recent years there has apparently been little new research activity directly on MT as such in Eastern Europe. In the Soviet Union it would appear that development of the AMPAR, NERPA and FRAP systems (cf. Hutchins 1986: 308-313) has not progressed significantly. The same seems to be true of MT research at Charles University Prague, which has always been a strong centre for computational linguistics (e.g. Hajičová and Sgall) and experimentation on advanced MT would seem likely in the future. In the meantime there is just reports of work on the dictionary formats for machine-aided English-Czech translation (Strossa 1987). Other relatively small projects are reported from Hungary and Bulgaria. At Szeged University (Hungary) Fabricz (1986) has been studying problems of modal particles (e.g. *only*, *nur*) in connection with an experimental English-Hungarian project. In Sofia (Bulgaria) a project for an English-Bulgarian system has been reported (Pericliev 1984, Pericliev & Iliaronov 1986) - the researchers argue that sentences which are ambiguous in the SL (English) do not need to be disambiguated if there exist parallel structures in the TL (Bulgarian) which maintain the same ambiguities. For example *'He promised to please mother*' and the corresponding Bulgarian *'Toj obešta da zaradva majka'*; in both cases the subordinated constructions (*to please mother*, *da zaradva majka*) can refer either to the object of the 'promise' or to the manner of the 'promise'. There are clear advantages in not having to disambiguate pronoun references if there is an equally 'ambiguous' pronoun in the TL. Further evidence of renewed interest in MT in Bulgaria was the holding of a conference in Sofia in May 1987 (LM48 Sept 1987).

25.

Without any doubt the country which has seen the most rapid expansion of MT research of all kinds is Japan. MT activity in Japan has been stimulated by a number of factors: the great demand for translations from English, in particular, of scientific and technical information, the demand for companies to produce English-language marketing and technical documentation for their products, the rapid growth of the Japanese export trade and of international competition, the great difficulties experienced by Japanese in learning European languages, the promotion of the Fifth Generation Project to give Japan a leading position in the future 'information society'. There are estimated to be now some 800-900 people presently engaged on research (Sigurdson & Greatrex 1987), with probably 60% of these in commercial companies, mainly involved in computer manufacture or in computer software.

The Japanese language has presented challenges to MT system design that were absent when most research concentrated on European languages. There are problems of the script: the combination of Japanese alphabets (hiragana and katakana) and Chinese characters, with no capitals and no spaces between words; and problems of language structure: a verb-final and modifier-modified language, no distinction of singular and plural nouns, no definite and indefinite articles, frequent omission of subjects (particularly pronouns), a largely free word order, numerous embedded compound clauses in sentences, much use of particles, and a 'logical' tense system. As a consequence, analysis of Japanese is oriented to the identification of semantic (case) relations rather than syntactic (phrase) structures, and operational systems often rely heavily on pre-editing (breaking up long sentences, inserting the omitted subjects, marking the modifying clauses, indicating the functions of particles). Most Japanese-English systems are intended for Japanese users who can do the pre-editing; Western users with no knowledge of Japanese will therefore get poor results, although the output may be good enough for information purposes (Sigurdson & Greatrex 1987).

Most systems and research projects are for translation between English and Japanese, although there is also interest in Japanese-Korean. General descriptions of current MT activity in Japan are to be found in Nishida & Doshita (1986), Whitelock (1987), Sigurdson & Greatrex (1987), and papers given at the MT Summit (1987).

26.

A pioneering centre for MT research has been Kyoto University under Makato Nagao, whose influence is evident in many current projects. The Japanese government MT project (the Mu project) at Kyoto was completed in March 1986 (Tsujii 1987). Its aim was to develop a prototype system for Japanese-English and English-Japanese translation for a restricted subject domain and document types. Mu is a bilingual transfer system, with dependency grammar analysis, a pre-transfer 'loop' to convert SL structures into more neutral forms, transfer proper (lexical and structural substitution), a post-transfer ‘loop’ for further adjustments to TL structures, and generation. Whereas analysis and generation of Japanese is wholly case grammar based, analysis and generation of English is partly based on a phrase structure grammar. An important contribution of the Kyoto research has been the development of a programming environment for grammar writing, the GRADE system (in conception similar to the 'special languages' of GETA.) A description of the Mu system is given by Nagao et al. (1985/1988, 1986), Nagao & Tsujii (1986), Nagao (1987). Evaluation of the output showed that unedited MT output was good enough for rough understanding of the gists of documents.

The Mu-II project is a four-year project (1986-1990) to transform the research prototype into a practical system for daily use by the Japan Information Center for Science and Technology (JICST) for translating abstracts. The project's aims are to reduce processing speeds and memory requirements, to enhance the dictionaries, to reduce the need for pre-editing, and to integrate text editing facilities. The development is being undertaken by researchers at Tokyo University, at the Electrotechnical Laboratory and at JICST, with the support of the Japanese government’s Science and Technology Agency. It has been reported that more than 50 researchers at JICST working full-time on MT: mostly on dictionaries, 10 on grammar and 10 on software (Sigurdson & Greatrex 1987). Sakamoto et al. (1986) describe the development of semantic markers, principally case frame features, to improve the quality of Japanese-English translation.

The other system developed at Kyoto, the restricted language system TITRAN for translating titles of scientific and technical papers, which was originally developed for English-Japanese, has now been extended to Japanese-English and Japanese-French. The English-Japanese version was evaluated and has been implemented in the Tsukuba Research Information Processing System of the Agency of Industrial Science and Technology. A collaborative project with Saarbrücken was also undertaken to adapt TITRAN to translating Japanese titles into German.

Since the ending of the research phase of Mu, the Kyoto MT team has been concentrating on basic theoretical work. Small scale projects have included research on a system written in Lisp for Chinese and Japanese, by Yang & Doshita (1986).

27.

There are numerous Japanese university research groups, e.g. at Tokyo Institute of Technology, Oita University, Kyushu University, Toyohashi University of Technology, Kobe University (where Sanamrad & Matsumoto (1986) have developed PERSIS, an analysis program for Persian). From the University of Tokyo Chung & Kunii (1986) have reported on the NARA system to translate from Korean into Japanese and vice versa; they claim that the syntactic structures of the two languages are sufficiently similar (both agglutinative and verb-final languages) for the formal transfer mechanism to be relatively simple. The project has concentrated so far on syntactic matters, adopting the formalism of Generalized Phrase Structure Grammar.

28.

The Japanese government is funding MT research through the Overseas Development Agency (ODA) and through its Key Technology Center (KTC), set up from capital resulting from the privatisation of NTT.

The Electronic Dictionary Research Institute (EDR), with 70% support from KTC and 30% from 8 electronics companies (Fujitsu, Hitachi, Mitsubishi. Matsushita, NEC, Oki, Sharp, Toshiba), is to collaborate with ICOT, the research facility established for the Fifth Generation Project (Sigurdson & Greatrex 1987). It will undertake basic research into dictionary systems and in the production of computer software programs for dictionary compilation. The project will develop two types of dictionaries (Kakizaki 1987). Word dictionaries, with entries under headwords giving definitions (concepts expressed) and grammatical features, will be developed both in English and in Japanese for 'basic' vocabulary (200,000 words of 'daily life') and for specialised terminology (100,000 words from the field of information processing). Concept dictionaries will be constructed using a 'knowledge representation language' which indicate possible (binary) relations between concepts: dependencies (e.g. cause-effect), synonymy, hypernymy, hyponymy, and thesaural relations by means of a concept classification. Major problems are recognised to be the maintenance of consistency, uniformity and accuracy. The dictionaries are to be evaluated by testing on MT systems, Information Retrieval systems and voice recognition systems. The main goal is the construction of dictionaries for interlingual MT systems (e.g. the ODA project, see below), and extension to other languages is envisaged.

Longer term research is being supported by KTC at Osaka, the seven-year project for Basic Research of Automatic Translation Telephone, funded 70% by KTC and 30% by the NTT and KDD telephone companies (Sigurdson & Greatrex 1987). The project has close links with the ATR (Advanced Telecommunications Research) project for automated telephony. At least 15 years basic research is anticipated; ATR originally offered collaboration with European and US companies (LM26 Nov 85). The approach at ATR has been particularly influenced by the MT research at NTT (see 36 below) by Nomura and his colleagues (e.g. Kogure & Nomura 1987).

The ODA project is an ambitious scheme to develop a multilingual system for translating between Japanese and languages of the Pacific economic region, Malay, Bahasa Indonesian, Thai and Chinese. It will be a collaborative project involving teams to be set up the countries concerned.

The six-year R&D project is receiving central funds from MITI (Ministry of International Trade and Industry) via CICC (Center of the International Cooperation for Computerization). The political aim is to promote technological and cultural exchange between Japan and other Asian countries, and to encourage the research capabilities of the countries concerned in the area of computer technology. Initial plans have been reported by Tsuji (1987); it will be an interlingua (text-analysis) system for two-way translation of industrial and technical information between Japanese and Thai, Japanese and Malay, Japanese and Indonesian, Japanese and Chinese (and with an eventual potential as a fully multilingual system supporting translation from and to Thai, Malay, Indonesian and Chinese). The aim is 80-90% accuracy with pre-editing, and a speed of 5000 words per hour. The project will concentrate on the development of general and special dictionaries, on the interlingua, on text analysis and generation procedures, and on a translation support system. In many respects the project is as ambitious in both political and linguistic terms as the Eurotra project (sect.12 above).

The research at the Japanese government Electrotechnical Laboratory (Tokyo) is now linked to the ODA project. It represents preliminary investigations of the ambitious text-based approach envisaged for the interlingua system. Ishizaki (1987) describes work on a small-scale experimental Japanese-English system using contextual information: CONTRAST (Context Translation System). Texts are interpreted ('understood') using information from a concept dictionary as well as grammatical and lexical information; in effect the concept dictionary will be the IL lexicon. The system will not adopt the Schank approach of semantic primitives and inference rules, but will attempt successive matching, reanalysis and conversion of SL input against concept structures. The results of 'contextual analysis' will be language-independent paragraph-based interface representations. The system has the following stages: syntactic analysis (augmented context free parser), semantic analysis, contextual analysis, paragraph-level generation, sentence-level generation, and word-level generation.

29.

At present, the operational MT systems in most common use in Japan are the batch-processing Systran system and the microcomputer-based Bravis system. However, they are rivalled now by other systems which have recently entered the market (29-34 below).

Systran Corporation of Japan (a company independent of the Systran companies owned by Gachot - cf.1 above) has developed English-Japanese and Japanese-English versions, running at speeds up to 2 million words per CPU hour on the FACOM M-380 (Akazawa 1986). Two major users of the English-Japanese system are a large Tokyo translation bureaux, which has been supplying post-edited Systran output since 1984 (over 10,000 pages per year), and the Tokyo branch of the Arthur Andersen accountancy company for the translation of its documentation into Japanese. The Japanese-English system has also been used outside Japan by the US Department of Defense and National Aeronautics and Space Administration and by the EEC for its Japan-Info project (Sigurdson & Greatrex 1987).

Bravis International (now owners of WCC, cf.6 above) has developed two versions of a Japanese-English system based on the Weidner design. The Minipack JE was introduced in 1984, running on DEC microcomputers under Unix, at a translation speed of 5-6000 words per hour. The Micro-Pack JE system was introduced in 1985, running under MS-DOS (640K memory, 20MB disk) at a translation speed of 1500 words per hour and integrating with word processor packages. Bravis is at present developing a MicroPack system for Japanese-Korean (Sigurdson & Greatrex 1987).

30.

The Fujitsu MT activity, which began in 1978, has taken place in the context of a wide range of long-term advanced research in the field of artificial intelligence (described by Sato &

Sugimoto 1986). Fujitsu has developed and now markets two systems: ATLAS/I for English-Japanese and ATLAS/II for Japanese-English. Current MT work also includes collaboration with the Korean Advanced Institute of Science and Technology on a Japanese-Korean system.

ATLAS/I, based on the 'direct translation' design (syntax-based, phrase structure grammar, coded in assembler) has been operational since 1982 (on test at Central Research Institute for Electrical Power Industries), and in 1984 appeared as the first commercial English-Japanese system (Sigurdson & Greatrex 1987).

ATLAS/II is a more advanced 'transfer' system for Japanese-English translation, written in the programming language C. It incorporates an AI-type 'world model' representing semantic relations between concepts for checking analysis and interface structures. SL and TL interfaces are conceptual dependency representations (case relations and semantic features), of sufficient abstractness to minimise transfer to conceptual and lexical differences as much as possible (i.e. many interface components are intended to be interlingual). The basic stages are as follows (Uchida 1986, 1987): morphological segmentation; simultaneous syntactic and semantic analysis (with a context-free grammar referring to the 'world model' for verification and disambiguation, checking of resulting SL semantic structure etc.); transfer of SL conceptual structure into TL conceptual structure; generation of TL linear form (dealing with syntactic and morphological restructuring simultaneously). The eventual aim is expansion into a multilingual system; there have been experiments in translating into French, German, Swahili, and Inuit: the most significant is the joint project with Stuttgart (SEMSYN) on Japanese-German (cf.16 above). ATLAS/II has been commercially available, within Japan only, since 1985, and has been ordered by EEC for the Japan-Info project to translate abstracts (Sigurdson & Greatrex 1987).

31.

The Hitachi company has also been involved for some years in two MT systems for Japanese-English and English-Japanese. There are now apparently over 100 people involved in MT, with perhaps 10-15 of these engaged on basic research (Sigurdson & Greatrex 1987).

The HICATS/JE (Hitachi Computer Aided Translation System/ Japanese to English) has been on the market since May 1987, in Japan only. Written in PL/I it is implemented on Hitachi M series computers (5MB memory, 100MB hard disk support) and is designed to handle scientific and technical documents, manuals and catalogues, with a processing speed up to 60,000 words an hour. The system was demonstrated at Geneva in 1987, via a satellite link (LT5 Jan 1988). Previously known as ATHENE/N (Nitta et al. 1984), the system is based on a transfer model with semantics-directed dependency structures as interface representations. The stages are: morphological analysis (segmentation into words); syntactic analysis (dependency grammar approach, using semantic features); semantic analysis (establishing case relations, performing disambiguation based on semantic features, producing conceptual dependency structure); transformation of dependency graph (conversion of SL-oriented representation to appropriate TL-oriented graph, and lexical substitution); syntactic generation (phrase grammar approach); morphological generation (Kaji 1987). The system provides for pre-editing, e.g. marking (bracketing) structurally ambiguous sections, and for post-editing (e.g. selection of alternative English words) in a split-screen interactive mode. It is supplied with a 250,000 base dictionary, and with facilities for users to compile their own dictionaries via a menu-driven editor.

The HICATS/EJ system for English-Japanese (also in PL/I and previously known as ATHENE/E) is not yet commercially available. In this case the transfer approach is more syntax-oriented, with the intention of adding semantic component to generation later. However, it is believed that because of the relatively free word order of Japanese stylistically awkward output may be acceptable - as long as the correct particles have been assigned. English analysis employs a phrase structure grammar, Japanese generation a case grammar approach.

Hitachi have recently (LM52 Jan 1988) announced the completion of a Japanese-Korean system developed in collaboration with the Korean Institute of Economics and Technology, with further plans for setting up a translation centre in Seoul and for developing a Korean-Japanese system.

32.

Toshiba's English-Japanese system AS-TRANSAC became available commercially during 1987, running at present only on Toshiba AS3000 minicomputers operating under Unix (8MB memory, 140MB disk support), with a claimed speed of 5-7000 words an hour (Sigurdson & Greatrex 1987). The Japanese-English version is still under development.

The English-Japanese system (known during its research phase as TAURUS, Toshiba Automatic Translation System Reinforced by Semantics) is a standard batch-processing bilingual transfer system, written in the C programming language. Its stages are as follows (Amano et al.1987): morphological analysis (assigning syntactic features); syntactic analysis (ATN parser producing a single purely syntactic representation, with all ambiguities remaining implicit); semantic analysis and transfer (using case frames and semantic features in an analysis oriented towards the requirements of Japanese, and also incorporating lexical transfer); syntactic generation (from Japanese conceptual representation, involving determination of word order and attachment of Japanese particles); and morphological generation.

Toshiba have demonstrated the potential of the two MT programs for international communication over telephone links (LT5 Jan 1988). The system is known as ATTP (=Automatic Translation Typing Phone). An English message is typed at a Toshiba AS-3000 engineering workstation, translated into Japanese by "station-resident software", and then transmitted via satellite; the reply is typed in Japanese at a similar workstation, translated into English, and "transmitted back almost simultaneously". The workstation displays messages sent and replies received on a split screen. ATTP is still a research prototype, with no date yet for a commercial product.

33.

The PENSEE systems from the Oki computer company are designed for Japanese-English and English-Japanese translation. The Japanese-English version has been commercially available since autumn 1986, running on Unix-based personal computers with 8MB memory and 80MB hard disk support, at a speed of 4000 words per hour. The English-Japanese version is still under development (Sigurdson & Greatrex 1987, MT Summit 1987).

PENSEE is described (Sakamoto & Shino 1987) as having these stages: morphological analysis, syntactic analysis, and English morphological synthesis. Analysis involves the identification of case relations and disambiguation using semantic features. The results are dependency (case) tree representations, which are then converted into English phrase structure representations by means of bilingual transfer dictionaries and tree transduction rules. It is basically the familiar syntactic transfer approach with limited semantics. The system is programmed in the C language, it includes facilities for bilingual editing and for in-house glossary compilation, and it provides for both sentence by sentence (i.e. 'interactive') and full text translation (i.e. batch). The designers acknowledged that translations are "literal" and "unnatural", and therefore need post-editing; also that some pre-editing may be necessary for poor Japanese input.

34.

The PIVOT system from NEC has been commercially available since 1986, running on an ACOS System 4 microcomputer (4MB memory, 120MB support disk), with a claimed maximum translation speed of 100,000 words per hour (Sigurdson & Greatrex 1987). Unlike other commercial Japanese systems, PIVOT (Pivot Oriented Translator) is based on the interlingua approach. Designed for bi-directional English and Japanese translation, the eventual aim is a multilingual system (Muraki 1987). Interlingua representations are in the form of networks of ‘conceptual primitives’ and 'case' labelled arcs, incorporating structural information, pragmatic information on topic and focus, theme and rheme, scopes of quantification and negation. In addition a knowledge base of extralinguistic information interacts at interface and interlingua levels for disambiguation. Written in the C programming language, processing passes through the stages of "morphological analysis, grammatical and semantic analysis, semantic extraction, conceptual wording, grammatical generation, morphological generation.” (MT Summit 1987). The problems of defining lexical items in terms of conceptual primitives are expected to be tackled within the EDR project on a conceptual dictionary (see 28 above). At present, PIVOT operates on dictionaries of 40,000 entries for Japanese and 53,000 entries for English, and covers some 20 different subject domains. In practical operation PIVOT can be used in batch mode or interactively. Facilities are provided for dictionary maintenance and for the essential processes of pre-editing and post-editing.

35.

The MELTRAN (Melcom Translation) from Mitsubishi is an interactive Japanese-English system which is to be available during 1988 (known during research as THALIA). It is a transfer system, with pre- and post-editing, either batch or interactive modes, with ‘logic programming’ in a version of Prolog (ESP), and running only on Melcom machines (10MB memory and 50MB disk support).

Ricoh 's RMT (Ricoh Machine Translation system), a transfer system for English-Japanese, is expected to be available commercially in 1988. Its features include English parsing with an augmented context free grammar, dependency trees with semantic features as interface representations, implementation in the C programming language, split-screen editing, OCR input, and a claimed translation speed of 4500 words per hour. Ricoh also has a Japanese-English system under development (MT Summit 1987, Sigurdson & Greatrex 1987).

The Sharp system OA-110WB for English-Japanese translation is also expected to reach the market in 1988. This is an interactive transfer system with pre- and post-editing essential, Written in C and operating under Unix, and running on Sharp machines OA-110WS, OA-210, OA-310 and IX-7, with a translation sped of 5000 words per hour. Analysis involves an augmented context free grammar, case frames, and some limited semantics. There is a basic dictionary of 60,000 words and technical dictionaries (up to 40,000 words) in the subject fields of economics, information processing, electronics, and mechanical engineering (MT Summit 1987, Sigurdson & Greatrex 1987).

The Sanyo system SWP-7800 system is advertised as a 'Translation Word Processor’ for Japanese-English translation, performing at a speed of 3500 words per hour. It became available in April 1987. It is basically a syntactic transfer model with pre- and post-editing, written in C and with augmented context free parsing and case grammar dependency structures. The basic dictionary contains 55,000 entries.

36.

The LUTE project at NTT (Nippon Telegraph and Telephone), which began in 1981, continues to represent one of the most advanced experimental systems. LUTE (Language Understander Translator Editor) is designed as an AI-oriented (knowledge-based) transfer system for bi-directional translation between Japanese and English. Recent accounts (Nomura et al. 1986; Kogura & Nomura 1987) describe the present configuration.

The basic formalism is a dynamic frame-network memory structure, comprising linguistic and non-linguistic knowledge, which is invoked by input text and which is used to interpret texts. As result of text processing and interpretation, the system assimilates the meaning of texts into its memory structure. Discourse structure, analogous to long-term memory is defined by Local Scene

Frame, a collection of cases and predicates of previously analysed sentences. Episodic (short-term) memory is modelled by the Extended Case Structure, a linguistic formalism for representing text meanings and for simultaneous representation of syntactic and semantic relations. It is not language-independent, thus there are separate ECS's for English and Japanese. The components of ECS are concepts (prototypes and instances), semantic categories, and semantic relations (whole-part, possession, case relations, modification, apposition, conjunction, time and cause relations, etc.) Therefore, the MT system analyses a Japanese sentence into a J-ECS, and its transfer process generates an equivalent E-ECS for English. Analysis ('extended case analysis') integrates semantic and syntactic information, intertwining bottom-up and top-down parsing. It includes establishment of syntactic patterns, semantic structures, case relations, modalities (passive, causative, etc.), resolution of ambiguities, and use of AI-type demons to perform specific tests, etc. Transfer processes are regarded as essentially frame manipulations. LUTE has monolingual dictionaries for analysis and generation (containing semantic categories, case frames, event relation frames, heuristics for resolving ambiguities, etc.), and a bilingual (concept) dictionary for transfer.

LUTE, implemented in ZetaLisp on a Symbolics Lisp machine, is undoubtedly one of the most sophisticated experimental projects at the present time; the researchers are not aiming to produce a prototype system but to provide a research environment for basic MT theoretical activity, in particular to investigate the relative roles of linguistic and non-linguistic knowledge at every stage of translation.

Another knowledge-based experimental transfer system is the LAMB system for Japanese-English translation under development by Canon. This is also written in Lisp, for a Symbolics 3620 machine. The PAROLE experimental system at Matsushita is another example: a Japanese-English transfer system with case-frame dependency structure, tree transducer, written in Prolog and Lisp (MT Summit 1987). And IBS KK are reported to be at any early stage in developing a bi-directional English and Japanese system (Sakurai 1987).

37.

Other companies have developed or are developing MT systems for internal use. These include the VALANTINE system at KDD; the prototype English-Japanese system at Nippon Data General Corporation (MT Summit 1987); and the English-Japanese system using a Lexical-Functional Grammar approach, developed by the CSK Research Institute (Kudo & Nomura 1986).

Best known, however, is IBM Japan's prototype English-Japanese system for translating IBM computer manuals (Tsutsumi 1986). The aim is a high-quality system within this particular restricted domain. The system is based on the transfer model, with the following stages: English analysis (using the augmented phrase structure grammar developed by Heidorn for the EPISTLE text-critiquing program); English transformation, in which some of the characteristically difficult structures of English are transformed into more Japanese-like structures (e.g. *'There are several records in the file'*→ *'Several records exist in the file'*); English-Japanese conversion (transfer proper) divided into semantic disambiguation (based on semantic markers appropriate for the domain of computer manuals), lexical conversion using 3 bilingual transfer dictionaries (for verbs, nouns and prepositions), substitution of expressions deemed to be idiomatic in this domain ('simple noun phrases'), and conversion of whole sentence structures (again employing semantic information); finally, Japanese generation.

38.

As we have seen, a number of Japanese projects involve Korean MT. Current research in Korea itself is reported by Kisik & Park (1987). As in Japan early efforts were devoted to the problems of character recognition, developing a Korean word processor, basic linguistic analysis of the language, and experimental application of different formalisms (case grammar, LFG, TG, etc.). Since the early 1980's MT research has been supported by the Korean Ministry of Science and

Technology at the Systems Engineering Research Institute (SERI) of the Korean Advanced Institute of Science and Technology (KAIST), and at four universities (Seoul, Inha, Hanyang, Graduate School of KAIST). The research at Seoul National University has concentrated on Korean-English transfer system (based on TG) and on a collaborative project with IBM on an English-Korean system. At Inha University there has been research on a prototype for bi-directional Korean and Japanese translation, involving semantic pattern matching and inference functions. The project at Hanyang University has been a LFG-based Japanese-Korean system in collaboration with Waseda University in Japan. The Graduate School of KAIST has undertaken fundamental research on computational parsers for Korean, experimenting with case grammar, Montague grammar, LFG and GPSG formalisms; recently a collaborative project with NEC has started on a multilingual system including Korean-Japanese and Korean-English. The SERI/KAIST project was the national Korean MT project for Korean and Japanese bi-directional translation based on a syntactic transfer model; the result has been KANT/I for Korean-Japanese running on a 16-bit Unix machine, and the cooperative development of a Japanese-Korean system with Fujitsu (see 30 above). SERI has also collaborated with GETA (see 13 above) on English-Korean and French-Korean systems using the ARIANE software. There is clear evidence of vigorous support for MT in Korean at a national level. The greatest demand is for translation between Korean and English, but for initial experiments Korean and Japanese offered more immediate expectations of success because of the affinity of the two languages. With the exception of the SERI/KAIST-Fujitsu project, few systems approach any kind of commercial feasibility.

38.

China's interest in MT can be dated back to the 1950's. A number of centres were active until the mid 1960's when internal political events brought them to an end for more than a decade. Recent activity began in 1979 with a cooperative project involving the Computational Institute of the Academy of Sciences, the Linguistic Institute of the Academy of Social Sciences and the Institute of Scientific and Technical Information. This was a project to translate titles of English metallurgical literature into Chinese. The most successful project so far has been KY-1, a transfer system developed at the Military Academy of Sciences. It is a post-edited batch transfer MT system (written in a derivative of COBOL) with some minimal pre-editing; analysis involves a "logical semantic theory" based on case grammar; an average accuracy of 75% is claimed. The system includes components for the compilation of English to Chinese dictionaries and for statistical analyses of English texts. It system is intended initially for translation in the fields of military science, electronics, chemistry and economics; it is to be marketed shortly as Transtar-1 and was demonstrated in December 1987 at a translation conference in Hong Kong (Kit-yee 1988; Du & Li 1987.

Other projects are an English-Chinese system under development at Tsinghua University of Beijing, a Chinese-English interactive system at the East China Normal University in Shanghai and an English-Chinese system at the National Tsing Hua University at Hsinchu (Taiwan). At Xi'an University researchers are developing a microcomputer-based English-Chinese system written in Prolog and operating in either batch or interactive modes. It is limited initially to the translation of catalogue titles. The system uses Wilks-type semantic templates to assign case frames and for disambiguating prepositions (Xi'an Univ. 1987).

Other non-university centres were mentioned by Dong Zheng Dong at the 1987 MT Summit conference. They include: the Institute of Linguistic Research, the Institute of Scientific and Technical Information, the Institute of Computer Software, the Academy of Posts and Telecommunications, the Academy of Military Sciences. Evidence of the growing interest in China are the establishment of a national committees for MT set up by Institute of Scientific and Technical Information, numerous regional and national conferences, and the international conferences organised by Chinese Information Processing Society in 1983 and 1987. A number of

Chinese scientists have been sent abroad to study MT, to Europe (particularly Grenoble) and to the United States (e.g. Texas).

40.

MT activity in other Asian countries has been growing rapidly. A joint project was started in 1979, involving the Grenoble group GETA and the University Sains Malaysia, to develop an English-Malay system based on the ARIANE-78 software (Vauquois and Boitet 1985/1988, Warotamasikkhadit 1986; Tong 1986). The initial aim was the translation of secondary-level teaching materials in technical fields and the basic system was completed in 1982. A test on a chemistry textbook in 1985 showed that 76% of sentences were ‘understandable’ without post-editing. Particular problems were English grammatical homonymy, prepositions, coordination, and Malay pronoun generation. In 1984 a permanent project was established to develop an industrial prototype within the next 3 years for texts in the field of computer science. The project will also be devoted to the development of a translator workstation for English-Malay translation, on the Melby model (cf.8 above), with automatic dictionary lookup, split-screen text processing, and eventual incorporation of the MT module (Tong 1987).

A similar cooperative project in Thailand involving the GETA group was set up in June 1981 using ARIANE software for an English-Thai system. This is being developed by researchers at the universities of Chulalongkorn, of Rakhamhaeng (Bangkok), and of Prince of Sonkia (Had-Yai); the project is being undertaken in cooperation with the English-Malay research at the University Sains Malaysia (Vauquois and Boitet 1985/1988, Warotamasikkhadit 1986).

The very early stages of MT activity in Indonesia have been reported by Sudarwo (1987). The national Agency for the Assessment and Application of Technology is at present developing a prototype machine-aided system for English-Indonesian, and is preparing the ground for involvement in the Japanese ODA multilingual project (cf.28 above).

In India research on MT began in 1983 at the Tamil University in South India. The TUMTS system is a small-scale 'direct translation' system specifically designed for Russian as SL and Tamil as TL and running on a small microcomputer with just 64K memory and restricted to a small corpus of astronomy text. The researchers readily admit the shortcomings and limitations of their work, but no doubt this modest start will encourage them and others in India.

41.

This 'geographical' survey ends with South America. The CADA project involving South American languages has been mentioned earlier (sect. 9). However, what has attracted much attention is the Bolivian MT project ATAMIRI. This is the system designed and developed by Ivan Guzman de Rojas (1985, 1986) of La Paz, Bolivia. It is a multilingual interlingua system using Aymara (an Indian language of South America) for intermediary representations (ATAMIRI is Aymaran for ‘translator’ and is an acronym for Automata Traductor Algoritmico Multilingue Interactivo Recursivo Inteligente). Guzman claims that the regularity of Aymara morphology and syntax make it ideal as an interlingua. After initial experiments with Spanish to Aymara translation Guzman began to investigate the use of Aymara as a 'pivot' language for multilingual translation of English, Spanish and German. Analysis and synthesis operate by matching structural patterns and transforming matrix representations; there is no use of tree structures. Updating of the dictionary is through the medium of Spanish. At present the system is purely syntactic, with no treatment of lexical and structural ambiguity. However, ATAMIRI is intended as an interactive system referring to professional translators for assistance. The system has been demonstrated at a number of places, including the World Monetary Fund and World Bank in Washington, and there are reports of its use by a translation centre in Panama for English-Spanish translation.

III. References

Akazawa, E. (1986): Systran Japanese systems. *Terminologie et Traduction* no.1, 1986, 78-89.

Alam, Y.S. (1986): A lexical-functional approach to Japanese for the purposes of machine translation. *Computers and Translation* 1(4), 1986, 199-214.

Amano, S., Hirakawa, H. and Tsutsumi, Y. (1987): TAURUS: the Toshiba machine translation system. In: MT Summit (1987), 15-23.

Ammon, R.von and Wessoly, R. (1984-85): Das Evaluationskonzept des automatischen Übersetzungsprojekts SUSY-DJT (Deutsch-Japanische Titelubersetzung). *Multilingua* 3(4),1984,189-195; & 4(1),1985,27-33.

Appelo, L. (1986): A compositional approach to the translation of temporal expressions in the Rosetta system. In: Coling 86, 313-318.

Appelo, L. and Landsbergen, J. (1986): The machine translation project Rosetta. Eindhoven: Philips Research Laboratories, 1986.

Arnold, D.J. (1986): Eurotra: a European perspective on MT. *Proceedings of the IEEE* 74(7), 1986, 979-992.

Arnold, D.J. et al. (1986): The <C,A>,T framework in Eurotra: a theoretically committed notation for MT. In: Coling 86, pp.297-303.

Arnold, D.J. and Des Tombe, L. (1987): Basic theory and methodology in Eurotra. In: Nirenburg (1987), 114-135.

Balfour, R.W. (1986): *Machine translation: a technology assessment*. London: BMT Consultants, 1986.

Barnes, J. (1987): User perspective on computer-assisted translation for minority languages. *Computers and Translation* 2(3), 1987,131-134

Batori, I. and Weber, H.J. eds. (1986): *Neue Ansätze in maschineller Sprachübersetzung: Wissensrepräsentation und Textbezug.* Tübingen: Niemeyer, 1986.

Bennett, P.A. et al. (1986): *Multilingual aspects of information technology*. Aldershot: Gower, 1986.

Bennett, W.S. and Slocum, J. (1985/1988): The LRC machine translation system. *Computational Linguistics* 11(2/3),1985, 111-121. Repr.in: Slocum (1988), 111-140.

Biewer, A. et al. (1985/1988): ASCOF: a modular multilevel system for French-German translation. *Computational Linguistics* 11(2/3), 1985, 137-155. Repr. in: Slocum (1988), 49-84.

Blatt, A., Freigang, K.H., Schmitz, K.D. and Thome, G. (1985): *Computer und Übersetzen: eine Einfuhrung*. Hildesheim: Olms, 1985.

Boitet, C. (1984/1987): Research and development on MT and related techniques at Grenoble University. In: King (1987), 133-153.

Boitet, C. and Gerber, R. (1986): Expert systems and other new techniques in M(a)T. In: Batori & Weber (1986), 103-119.

Boitet, C. (1986): The French national MT-project: technical organization and translation results of CALLIOPE-AERO. *Computers and Translation* 1(4), 1986, 239-267.

Boitet, C. (1987a): Current state and future outlook of the research at GETA. In: MT Summit (1987), 26-35.

Boitet, C. (1987b): Current projects at GETA on or about machine translation. Presented at II World Basque Congress, San Sebastian, September 1987.

Bostad, D.A. (1985): Soviet patent bulletin processing: a particular application of machine translation. *CALICO Journal* 2(4), June 1985, 27-30.

Bostad, D. (1986): Machine translation in the USAF. *Terminologie et Traduction* no.1, 1986, 68-72.

Brekke, M. and Skarsten, R. (1987): Machine translation: a threat or a promise? Paper presented at Translating and the Computer 9, November 1987.

Buchmann, B., Warwick, S. and Shann, P. (1984): Design of a machine translation system for a sublanguage. In: Coling 84, 334-337.

Carbonell, J.G. and Tomita, M. (1987): Knowledge-based machine translation, the CMU approach. In: Nirenburg (1987), 68-89.

Chauché, J. (1986): Deduction automatique et systèmes transformationnels. In: Coling 86, 408-411.

Chellamuthu, K.C., Rangan, K. and Murugesan, K.C. (1984): *Tamil University Machine Translation System (TUMTS), Russian-Tamil.* Research report, phase-I. Thanjavur: Tamil Univ., 1986.

Chung, H.S. and Kunii, T.L. (1986): NARA: a two-way simultaneous interpretation system between Korean and Japanese. In: Coling '86, pp.325-328.

Coling 84. 10th International Conference on Computational Linguistics, 22nd Annual meeting of the Association for Computational Linguistics: proceedings of Coling84. Stanford: University, 1984.

Coling 86. 11th International Conference on Computational Linguistics: proceedings of Coling '86. Bonn: University, 1986.

Darke, D. (1986): Machine translation for Arabic. *Language Monthly* 28, 1986, 10-11.

De Jong, F. and Appelo, L. (1987): *Synonymy and translation*. Eindhoven: Philips Research Laboratories, 1987.

Dowty, D.R., Wall, R.E. and Peters, S. (1981): *Introduction to Montague semantics*. Dordrecht: Reidel, 1981.

Du, C.Z. and Li, R. (1987): *An overview of MT research in China*. Xi'an University, 1987.

Ducrot, J.M. (1985): TITUS IV: System zur automatischen und gleichzeitigen Übersetzung in vier Sprachen. *Sprache und Datenverarbeitung* 9(1), 1985, 28-36.

Fabricz, K. (1986): Particle homonymy and machine translation. In: Coling '86, pp.59-61.

Gazdar, G. et al. (1985): *Generalized phrase structure grammar*. Oxford: Blackwell, 1985.

Goshawke, W. et al. (1987): *Computer translation of natural language*. Wilmslow: Sigma Press, 1987

Guilbaud, J.P. (1984/1987): Principles and results of a German to French MT system at Grenoble University (GETA). In: King (1987), 278-318.

Guzman de Rojas, I. (1985): Hacia una ingenieria del lenguaje. *Boletin de Informatica RCII* 3, 1985, 1-10 & 18-25.

Guzman de Rojas, I. (1986): Seminario sobre traduccion multilingue por computadora utilizando el sistema ATAMIRI, Roma, noviembre de 1986. Roma: Intergovernmental Bureau for Informatics, 1986.

Habermann, F.W.A. (1986): Provision and use of raw machine translation. *Terminologie et Traduction* no.1, 1986, 29-42.

Hauenschild, C. (1986): KIT/NASEV oder die Problematik des Transfers bei der maschinellen Übersetzung. In: Batori & Weber (1986), 167-195.

Hayes, P.J. (1984): Entity-oriented parsing. In: Coling 84, 212-217.

Huang, X. (1986): Processing in parallel. In: Coling '86, 140-145.

Hutchins, W.J. (1986): *Machine translation: past, present, future*. Chichester: Ellis Horwood, 1986. (New York: Halstead, 1986)

Hutchins, W.J. (1987): Prospects in machine translation. In: MT Summit (1987), 48-52.

Isabelle, P. and Bourbeau, L. (1985/1988): TAUM-AVIATION: its technical features and some experimental results. *Computational Linguistics* 11(1), 1985, 18-27. Repr. in: Slocum (1988), 237-263.

Isabelle, P. and Macklovitch, E. (1986): Transfer and MT modularity. In: Coling '86, pp. 115-117.

Ishizaki, S. (1987): Machine translation using contextual information. In: MT Summit (1987), 53-54.

Jin, W. and Simmons, R.F. (1986): Symmetric rules for translation of English and Chinese. *Computers and Translation* 1(3), 1986, 153-167

Johnson, R.L., Krauwer, S., Rosner, M. and Varile, G.B. (1984): The design of the kernel architecture for the Eurotra software. In: Coling 84, 226-235.

Johnson, R.L., King, M. and des Tombe, L. (1985): Eurotra: a multi-lingual system under development. *Computational Linguistics* 11(2/3), 1985, 155-169.

Johnson, R.L. and Whitelock, P. (1987): Machine translation as an expert task. In: Nirenburg (1987), 136-144.

Johnson, T. (1985): *Natural language computing: the commercial applications*. London: Ovum Ltd., 1985.

Joscelyne, A. (1987): Calliope and other pipe dreams. *Language Technology* 4 (Nov/Dec 1987), 20-21.

Joscelyne, A. (1988): Jean Gachot resurrecting Systran. *Language Technology* 6 (March/April 1988), 26-29.

Kaji, H. (1987): HICATS/JE: a Japanese-to-English machine translation system based on semantics. In: MT Summit (1987), 55-60.

Kakizaki, N. (1987): Research and development of an electronic dictionary: current status and future plan. In: MT Summit (1987), 61-64.

Kaplan, R.M. and Bresnan, J. (1983): Lexical-functional grammar: aformal system for grammatical representations. In: Bresnan, J. (ed.): *The mental representation of grammatical relations*. Cambridge: MIT Press, 1983.

Kasper, R. and Weber, D. (1986): *User's reference manual for the C Quechua Adaptation Program*. Dallas, TX: Summer Institute of Linguistics, 1986.

Kay, M. (1984): Functional unification grammar: a formalism for machine translation. In: Coling 84, 75-78.

King, M. ed. (1983): *Parsing natural language*. London: Academic Press, 1983.

King, M. ed. (1987): *Machine translation today: the state of the art*. Proceedings of the Third Lugano Tutorial 2-7 April 1984. Edinburgh: Edinburgh Univ.Press, 1987.

Kisik, L. and Park, C. (1987): The machine translation researches and governmental view in Korea. In: MT Summit (1987), 66-72.

Kit-yee, P.C. (1988): The future of Chinese translation. *Language Monthly* 53 (February 1988), 6-10.

Knowles, F.E. (1987): The Alvey Japanese and English machine translation project. In: MT Summit (1987), 73-78.

Kogure, K. and Nomura, H. (1987): Computer environment for meaning structure representation and manipulation in machine translation system. Paper presented at International Conference on Information and Knowledge, November 1987, Yokohama.

Kudo, I. and Nomura, H. (1986): Lexical-functional transfer: a tranfer framework in a machine translation system based on LFG. In: Coling '86, pp.112-114.

Landsbergen, J. (1984/1987): Isomorphic grammars and their use in the Rosetta translation system. In: King (1987), 351-372.

Landsbergen, J. (1987): Montague grammar and machine translation. In: Whitelock, P. et al. (eds.) *Linguistic theory and computer applications* (London: Academic Press, 1987), 113-147

Laubsch, J. et al. (1986): Language generation from conceptual structure: synthesis of German in a Japanese/ German MT project. In: Coling 84, 491-494.

Lee, C. (1987): Machine translation: English-Korean. In: *The Thirteenth LACUS Forum 1986*, ed. I.Fleming (Lake Bluff, Ill.: Linguistic Association of Canada and the United States, 1987), 410-416.

Leermakers, R. and Rous, J. (1986): The translation method of Rosetta. *Computers and Translation* 1(3), 1986, 169-183.

Lehrberger, J. and Bourbeau, L. (1988): *Machine translation: linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam: Benjamins, 1988.

Leon, M. and Schwartz, L.A. (1986): *Integrated development of English-Spanish machine translation: from pilot to full operational capability*. Washington,D.C.: Pan American Health Organization, 1986.

Lewis, D. (1985): The development and progress of machine translation systems. *ALLC Journal* 5, 1985, 40-52.

Licher, V., Luckhardt, H.D. and Thiel, M. (1987): Konzeption, computerlinguistische Grundlagen und Implementierung eines sprachverarbeitenden Systems. In: Wilss & Schmitz (1987), 113-153.

Liu, J. and Liro, J. (1987): The METAL English-to-German system: first progress report. *Computers and Translation* 2(4), 1987, 205-218.

Loomis, T. (1987): Software design issues for natural language processing. *Computers and Translation* 2(4), 1987, 219-230.

Luckhardt, H.D. (1987a): *Der Transfer in der maschinellen Sprachübersetzung* (Sprache und Information 18). Tübingen: Niemeyer, 1987.

Luckhardt, H.D. (1987b): *Von der Forschung zur Anwendung: das computergestützte Saarbrücker Translationssystem STS*. Saarbrücken: Univ.d.Saarlandes, 1987.

Luctkens, E. and Fermont, P. (1986): A prototype machine translation based on extracts from data processing manuals. In: Coling '86, pp.643-645.

Maas, H.D. (1984/1987): The MT system SUSY. In: King (1987), 209-246.

McDonald, D.D. (1987): Natural language generation: complexities and techniques. In: Nirenburg (1987), 192-224.

Mann, J.S. (1987): Get Smart! industrial strength language processing from Smart Communications. *Language Technology* 3 (September/October 1987),12-15.

Melby, A.K. (1986): Lexical transfer: a missing element in linguistics theories. In: Coling '86, pp.104-106.

Moore, G.W. et al. (1986): Automated translation of German to English medical text. *American Journal of Medicine* 81, 1986, 103-111.

MT Summit (1987): Machine Translation Summit manuscripts & program, September 17-19, 1987, Hakone, Japan.

Muraki, K., Ichiyama, S. and Fukumochi, Y. (1985): Augmented dependency grammar: a simple interface between the grammar rule and the knowledge. In: *Second Conference of the European Chapter of the Association for Computational Linguistics*... March 1985, University of Geneva, 198-204.

Muraki, K. (1987): PIVOT: two-phase machine translation system. In: MT Summit (1987), 81-83.

Nagao, M. (1985/1987): Role of structural transformation in a machine translation system. In: Nirenburg (1987), 262-277.

Nagao, M., Tsujii, J.I. and Nakamura, J.I. (1985/1988): The Japanese government project for machine translation. *Computational Linguistics* 11(2/3), 1985, 91-110. Repr.in: Slocum (1988), 141-186.

Nagao, M., Tsujii, J.I. and Nakamura, J.I. (1986): Machine translation from Japanese into English. *Proceedings of the IEEE* 74(7), 1986, 993-1012.

Nagao, M. and Tsujii, J.I. (1986): The transfer phase of the Mu machine translation system. In: Coling 86, 97-103.

Nirenburg, S. ed. (1987): *Machine translation: theoretical and methodological issues*. Cambridge: Cambridge Univ.Press, 1987.

Nirenburg, S., Raskin, V. and Tucker, A.B. (1986): On knowledge-based machine translation. In: Coling '86, 627-632.

Nirenburg, S., Raskin, V. and Tucker, A.B. (1987): The structure of interlingua in TRANSLATOR. In: Nirenburg (1987), 90-113.

Nirenburg, S. and Carbonell, J. (1987): Integrating discourse pragmatics and propositional knowledge for multilingual natural language processing. *Computers and Translation* 2(2), 1987, 105-116.

Nishida, T. and Doshita, S. (1986): Machine translation: Japanese perspectives. In: Picken (1986), 152-174.

Nitta, Y. et al. (1984): A proper treatment of syntax and semantics in machine translation. In: Coling 84, 159-166.

Papegaaij, B.C. (1986): *Word expert semantics: an interlingual knowledge-based approach* (Distributed Language Translation 1). Dordrecht: Foris, 1986.

Pericliev, V. (1984): Handling syntactical ambiguity in machine translation. In: Coling 84, 521-524.

Pericliev, V. and Ilarionov, I. (1986): Testing the projectivity hypothesis. In: Coling '86, pp.56-58.

Picken, C. ed. (1985): *Translation and communication: Translating and the computer 6*. Proceedings of a conference... November 1984. London: Aslib, 1985.

Picken, C. ed. (1986): *Translating and the computer 7*. Proceedings of a conference... November 1985. London: Aslib, 1986.

Picken, C. ed. (1987): *Translating and the computer 8: a profession on the move*. Proceedings of a conference... November 1986. London: Aslib, 1987.

Pigott, I.M. (1986): Essential requirements for a large-scale operational machine-translation system. *Computers and Translation* 1(2), 1986, 67-72

Pigott, I.M. (1988): Systran machine translation at the EC Commission: present status and history. [Luxembourg: CEC, January 1988.]

Popesco, L. (1986): Limited context semantic translation from a single knowledge-base for a natural language and structuring metarules. *Computers and the Humanties* 20, 1986, 289-295.

Reed, R.B. (1985): CADA: an overview of the Tucanoan experiment. *Notes on Computing (SIL)* 11, 1985, 6-20.

Roesner, D. (1986a): SEMSYN - Wissensquellen und Strategien bei der Generierung von Deutsch aus einer semantischen Repräsentation. In: Batori & Weber (1986), 121-137.

Roesner, D. (1986b): When Mariko talks to Siegfried. In: Coling 86, 652-654.

Rohrer, C. (1986a): Maschinelle Übersetzung mit Unifikationsgrammatiken. In: Batori & Weber (1986), 75-99.

Rohrer, C. (1986b): Linguistic bases for machine translation. In: Coling '86, 353-355.
Rolf, P.C. and Chauche, J. (1986): Machine translation and the SYGMART system. *Computers and the Humanities* 20, 1986, 283-288.
Rothkegel, A. (1986a): Textverstehen und Transfer in der maschinellen Übersetzung. In: Batori & Weber (1986), 197-227.
Rothkegel, A. (1986b): Pragmatics in machine translation. In: Coling '86, 335-337.
Rothkegel, A. (1987): Semantisch-pragmatische Aspekte in der maschinellen Übersetzung. In: Wilss & Schmitz (1987), 163-180.
Ryan, J.P. (1987): SYSTRAN: a machine translation system to meet user needs. In: MT Summit (1987), 99-103.
Sadler, V. and Papegaaij, B.C. (1987): The Melby test on knowledge-based lexical transfer. Utrecht: BSO, 1987.
Sakamoto, M. and Shino, T. (1987): Japanese-English machine translation system 'PENSEE'. *Oki Technical Review* 126, April 1987, 9-14.
Sakamoto, Y. et al. (1986): Concept and structure of semantic markers for machine translation in Mu-project. In: Coling '86, pp.13-19.
Sakurai, K. (1987): On machine translation. *Information, Information Processing, Information Retrieval/Joho Kagaku* 23(4), 178-181. (In Japanese; Japan Technology 0119542)
Sanamrad, M.A. and Matsumoto, H. (1986): PERSIS: a natural language analyzer for Persian. *Journal of Information Processing* 8(4), 1986, 271-279.
Sato, S. and Sugimoto, M. (1986): Artificial intelligence. *Fujitsu Science and Technology Journal* 22(3), 1986, 139-181.
Scheel, H.L. (1987): Bericht uber die Arbeit des Projektbereichs C. In: Wilss & Schmitz (1987), 211-223.
Schenk, A. (1986): Idioms in the Rosetta machine translation system. In: Coling 86, 319-324.
Schmidt, P. (1986): Valency theory in a stratificational MT-system. In: Coling '86, 307-312.
Schneider, T. (1987): The METAL system, status 1987. In: MT Summit (1987), 105-112.
Schubert, K. (1986): Linguistic and extra-linguistic knowledge. *Computers and Translation* 1(3), 1986, 125-152.
Schubert, K. (1987): *Metataxis: contrastive depenedency syntax for machine translation*. (Distributed Language Translation 2). Dordrecht: Foris, 1987.
Siebenaler, L. (1986): Systran for ESPRIT and ECAT bureau service. *Terminologie et Traduction* no.1, 1986, 54-60.
Sharp, R. (1986): A parametric NL translator. In: Coling '86,pp.124-126.
Shieber, S.M. (1986): *An introduction to unification-based approaches to grammar*. Stanford, Ca: Center for the Study of Language and Information, 1986.
Sigurdson, J. and Greatrex, R. (1987): *Machine translation of on-line searches in Japanese databases.* Lund: Research Policy Institute, Lund Univ., 1987.
Slocum, J. (1985/1988): A survey of machine translation: its history, current status, and future prospects. *Computational Linguistics* 11(1), 1985, 1-17. Repr.in: Slocum (1988), 1-47.
Slocum, J. (1987): Concept-lexeme-syntax triangles: a gateway to interlingual translation. *Computers and Translation* 2(4), 1987, 243-261.
Slocum, J. ed. (1988): *Machine translation systems*. Cambridge: Cambridge Univ.Press, 1988.
Smith, D. (1987): Translation practice in Europe. In: Picken (1987), 74-82.
Somers, H.L. (1986a):The need for MT-oriented versions of case and valency in MT. In: Coling '86, 118-123.
Somers, H.L. ed. (1986): Eurotra special issue. *Multilingua* 5(3), 1986, 129-177.
Somers, H.L. (1987a): Some thoughts on interface structure(s). In: Wilss & Schmitz (1987), 81-99.
Somers, H.L. (1987b): *Valency and case in computational linguistics*. Edinburgh: Edinburgh Univ.Press, 1987.
Stegentritt, E. (1987): Überblick und Bericht zum ASCOF-System des Projekts C. In: Wilss & Schmitz (1987), 225-231.
Steiner, E. (1986): Generating semantic structures in Eurotra-D. In: Coling '86, 304-306.
Stentiford, F. and Steer, M.G. (1987): A speech driven language translation system. Presented at European Speech Technology Conference, Edinburgh, September 1987.

Strossa, P. (1987): SPSS - an algorithm and data structures design for a machine aided English-to-Czech translation system. *Prague Bulletin of Mathematical Linguistics* 47, 1987, 25-36.

Sudarwo, I. (1987): The needs of MT for Indonesia. In: MT Summit (1987), p. 113.

Tomita, M. (1984): Disambiguating grammatically ambiguous sentences by asking. In: Coling 84, 476-480.

Tomita, M. (1986): Sentence disambiguation by asking. *Computers and Translation* 1(1), 1986, 39-51

Tomita, M. and Carbonell, J.G. (1986): Another stride towards knowledge-based machine translation. In: Coling '86, 633-638.

Tong, L-C. (1986): English-Malay translation system: a laboratory prototype. In: Coling '86, pp.639-642.

Tong, L-C. (1987): The engineering of a translator workstation. *Computers and Translation* 2(4), 1987, 263-273.

Trabulsi, S. (1986): Difficultes de la traduction en langue arabe. *Terminologie et Traduction* no.1, 1986, 90-95.

Tsuji, Y. (1987): Research and development project of machine translation system with Japan's neighboring countries. In: MT Summit (1987), 116-120.

Tsujii, J.I. (1986): Future directions of machine translation. In: Coling '86, 655-668.

Tsujii, J.I. (1987): The current stage of the Mu-project. In: MT Summit (1987), 122-127.

Tsutsumi, T. (1986): A prototype English-Japanese machine translation system for translating IBM computer manuals. In: Coling '86, 646-648.

Tucker, A.B. (1987): Current strategies in machine translation research and development. In: Nirenburg (1987), 22-41.

Uchida, H. (1986): Fujitsu machine translation system ATLAS. *Future Generations Computer Systems* 2, 1986, 95-100.

Uchida, H. (1987): ATLAS: Fujitsu machine translation system. In: MT Summit (1987), 129-134.

Vasconcellos, M. and Leon, M. (1985/1988): SPANAM and ENGSPAN: machine translation at the Pan American Health Organization. *Computational Linguistics* 11(2/3), 1985, 122-136. Repr. in Slocum (1988), 187-235.

Vasconcellos, M. (1986): Functional considerations in the postediting of machine translated output. I.Dealing with V(S)O versus SVO. *Computers and Translation* 1(1), 1986, 21-38

Vasconcellos, M. (1987): Post-editing on-screen: machine translation from Spanish into English. In: Picken (1987), 133-146.

Vauquois, B. and Boitet, C. (1985/1988): Automated translation at Grenoble University. *Computational Linguistics* 11(1), 1985, 28-36. Repr. in Slocum (1988), 85-110.

Warotamasikkhadit, U. (1986): Computer aided translation project, University Sains Malaysia, Penang, Malaysia. *Computers and Translation* 1(2), 1986, 113.

Weber, H.J. (1986): Faktoren einer textbezogenen maschinellen Übersetzung: Satzstrukturen, Kohärenz- und Koreferenz-Relationen, Textorganisation. In: Batori & Weber (1986), 229-261.

Weber, H.J. (1987): Wissensrepräsentation und Textstruktur-Analyse. In: Wilss & Schmitz (1987), 181-209.

Wheeler, P. (1985): LOGOS. *Sprache und Datenverarbeitung* 9(1), 1985, 11-21.

White, J.S. (1987): The research environment in the METAL project. In: Nirenburg (1987), 225-246.

Whitelock, P.J. et al. (1986): Strategies for interactive machine translation: the experience and implications of the UMIST Japanese project. In: Coling '86, pp.329-334.

Whitelock, P.J. (1987): Japanese machine translation in Japan and the rest of the world. in: Picken (1987), 147-159.

Wilss, W. and Schmitz, K.D. eds. (1987): *Maschinelle Übersetzung, Methoden und Werkzeuge*: Akten des 3.Internationalen Kolloquiums des Sonderforschungsbereichs 100... Saarbrücken... September 1986. Tübingen: Niemeyer, 1987.

World Systran Conference (1986). *Terminologie et Traduction* 1986, no.1 special issue.

Xi'an University (1987): Resarch of machine translation on microcomputer, by MT Group Xi'an Jiaotong University.

Yang, Y. and Doshita, S. (1986): The construction of a semantic grammar for Chinese language processing and its implementation. *Transactions of the Information Processing Society of Japan* 27(2), 1986, 155-164. (In Japanese: Japan Technology 0049845)

Yoshii, R. (1986): A robust machine translation system. *Proceedings of SPIE 635*, Applications of Artificial Intelligence III, 1986, 455-463

Zajac, R. (1986): SCSL: a linguistic specification language for MT. In: Coling86, pp.393-398.

Zimmermann, H.H., Kroupa, E. and Luckhardt, H.D. eds. (1987): *Das Saarbrücker Translationssystem STS:* eine Konzeption zur computergestützten Übersetzung. Saarbrücken: Univ.d.Saarlandes, 1987.