Chapter 17: Interactive and human-aided systems.

The experience of the last decades has revealed the strengths and weaknesses of computers in handling natural languages and in translating from one language into another. Dictionary lookup and morphological analysis are now largely unproblematic; but certain types of syntactic ambiguity and wide areas of semantic analysis continue to cause problems. Many of these problems appear to be intractable, or at least to be still far from satisfactory solution or adequate treatment. However, there is now a sufficient body of well-tested techniques for the construction of economically viable practical systems by combining reliable computational methods and human expertise. The systems remain within the presently known limitations of the computer. There are two basic approaches: either input is deliberately restricted, or the expertise of human translators is called upon during translation processes. In the first case, texts are simplified for the system or the system itself is simplified; in the second case, translations are produced interactively by exploiting the on-line text-editing facilities of modern computer equipment.

17.1: Pre-edited input.

Proposals for simplifying or doctoring MT input texts were made in the earliest days of MT activity. Reifler suggested that a 'pre-editor' (Ch.2.4.1) could mark up SL texts so that analysis could be made easier. Dodd proposed that texts for MT could be written in a simplified regularised English (Ch.2.4.3). More recently the idea has appeared again in organisations which, while using a MT system otherwise capable of accepting unedited input, limit the vocabulary and grammar of texts for translation in order to reduce problems of post-editing.

A number of companies which have installed operational systems have established standards for SL text input. The Xerox Company (Elliston 1979, Ruffino 1982) installed the Systran system (Ch.12.1) for translating technical and servicing manuals. An initial experiment with uncontrolled input resulted in texts requiring an unacceptable amount of post-editing. It was decided to experiment with control of the input. The company had previously had experience with "Caterpillar English", a highly limited version of English similar to Ogden's Basic English (1968). This had proved unacceptable to personnel already used to documentation of a high standard. For its own controlled English, the Xerox Company has developed 'Multinational Customized English' (MCE) based on a general and technical vocabulary of 3000 words and rules for writing unambiguous English. For example, *replace* may express two different actions, as in*: Remove part A and replace it with part B*. This is clearer as: *Remove part A, adjust part B and then replace part A*. It was found that such steps not only improved Systran performance but also improved the English of the authors of technical manuals. Fears that MCE would be a kind of pidgin English were unfounded. Systran translations with MCE input are being produced from English into French, Spanish and Italian.

17. 2: Restricted language systems

Proposals for simplifying the MT system itself are based on the recognition that in certain circumstances the grammar and vocabulary of input texts constitute an identifiable and definable sublanguage. The MT system can then be designed specifically for such 'restricted language' texts. One example has already been described, the successful METEO system for translating meteorological reports into French (Ch.13.1.2). This system was devised by the TAUM team which pioneered the study of sublanguages, particularly Kittredge and Lehrberger (1982).

17. 3: Institut Textile de France: TITUS system (1970-

TITUS is a system devised specifically for input which is highly restricted in both vocabulary and syntax. TITUS was introduced by the Institut Textile de France in 1970 for the translation of abstracts from and into four languages: French, English, German and Spanish (Ducrot 1973) Abstracts for TITUS are formulated as sequences of clauses consisting of only permitted keywords (index terms) and a fixed set of 'function words' (prepositions, conjunctions, negatives, auxiliary verbs and modals). Nouns may be compound, verbs may be inflected, and adjectives may be simple, compound, or verbal derivatives. Abstracts are input and checked on-line, with interactive prompting of abstractors for reformulations if the TITUS vocabulary or syntax are violated. Abstracts are stored in a 'swivel language' (i.e. interlingua) from which they are translated into one or more of the four languages. The TITUS syntax permits a sequence of up to four clauses: main clause followed by subordinate or relative clauses. Each clause consists of the sequence: subject, prepositional phrase, verbal group, complement of verb, and prepositional phrase. In order to avoid the confusion of subordinating and prepositional uses of *before*, *after*, *since*, etc. diacritical marks are added to indicate the beginnings of subordinate clauses. The latest version TITUS 4 was implemented in 1980, and is claimed to permit abstracts to be written in "sentences or phrases which closely resemble free-language sentences or phrases" (Streiff 1985).[1]

17. 4: SMART AI, Inc. (1977-

In the belief that linguistic approaches to MT were misguided, John M. Smart based a computer-aided system on the premise that MT was possible only if the vocabulary was well defined. The foundation was provided by Caterpillar Fundamental English (which had attracted Xerox); the system was developed by Smart Communications Inc. (now Smart AI Inc.) into a computer-aided translation system, programmed in Burroughs Algol, and marketed first in 1978 (Smart 1985).

The Smart system was adopted by the Canadian Department of Employment and Immigration for translating job descriptions from English into French. These are limited in syntax (basically sequences of noun phrases) and with standardised vocabularies. The initial trials began in February 1982, and the full system has been operative in the Toronto region since early 1983. The dictionaries have now reached about 10,000 entries. Smart claims that the system can deal with more complex input: technical manuals, telex messages, engineering specifications. Evidently, there are clearly considerable constraints in input vocabulary[2]. Details of the linguistic processes of the system are not available; it is claimed to be an "expert system for translation" (Smart 1985), but to what extent AI techniques are employed is not known.

17. 5: Other 'restricted' systems

The most obvious application of restricted language MT is in the translation of titles. The TITRAN system developed at Kyoto University (Ch.18.13 below) has been devised to translate English titles of articles in the field of metallurgy into Japanese. Another system recently reported (Valle Bracero & Fernández García 1982) translates the titles of Russian journal articles into Spanish. The titles are taken from the metallurgy sections of the *Referativnyi Zhurnal*. The system has been developed at the Instituto de Informacion y Documentacion en Ciencia y Tecnologia in Madrid.

Finally, there has been a proposal to use a version of Basic English in SL texts by Straub and Rogers (1979). The experimental BEAST system (Basic English Analyser System), written in

---

[1] For a full description see also J.M.Ducrot 'Le système TITUS IV: système de traduction automatique et simultanée en quatre langues', *La traduction assistée par ordinateur*, ed. A. Abbou (Paris: DAICADIF, 1989), pp.55-67

[2] Subsequent accounts confirm that the Smart Translator is designed specifically for controlled language input, using the Smart Expert Editor (MAX, later MAXit) – see J.Hutchins 'Recent developments in machine translation: a review of the last five years', *New directions in machine translation: international conference, Budapest 18-19 August 1988*, ed. D.Maxwell, K.Schubert, T.Witkam (Dordrecht: Foris, 1988), 7-63; and A. Lee 'Controlled English with and without machine translation', *Machine translation today: Translating and the Computer 15. Papers presented at a conference... 18-19 November 1993... London* (London: Aslib, 1993), 35-39

COBOL, was designed to demonstrate the feasibility of a SL analysis program. The intention was that analysis would be interactive (as in the Brigham Young system, Ch.17.10) and output would be in Basic versions of other languages. Of a similar nature also was the proposal by Goshawke (1977) for a numerical interlingua, SLUNT (Spoken Languages Universal Numeric Translation), which was to operate on simplified SL (English) input.

17. 6: Automated dictionaries and term banks.

Advances in computer technology since the early 1970s have made possible other conceptions of how man and machine may collaborate in translation processes. The crucial developments are time-sharing systems, on-line access, text editing facilities, and remote access over telecommunications networks.

One of the earliest implementations was the project by Erhard Lippmann at the IBM Research Center, Yorktown Heights, N.Y. He investigated an experimental computer-aided translation system, designed as an 'extension' of the translator's capabilities (Lippmann 1971). Seated at a terminal, the translator would be able to "1) enter and/or edit a text, e.g. a translation or a dictionary; 2) look up words or phrases in a dictionary, 3) update dictionaries or other text files, 4) print his text volumes completely or selectively..." etc. In other words, Lippmann was describing what is now known as a 'word processor' used in combination with automatic dictionaries or terminology databanks.

Machine aids for translators are now commonplace in many translation offices and service bureau. Many translators and translation offices create their own in-house dictionaries, not only because published dictionaries in scientific and technical fields are inevitably out of date but in order to ensure consistent usage. Larger translation services, such as the European Communities in Brussels, have made their databanks available for other users. There are now numerous multilingual and monolingual terminology databanks for specialised scientific, technical, administrative, economic, political, agricultural, medical, etc. vocabulary, e.g. EURODICAUTOM at Brussels, LEXIS at the German Bundessprachenamt in Bonn, TERMIUM at Montreal, and TEAM at Siemens, to mention only the best known.

It should be noted that the information contained in computer-based dictionaries does not coincide with that contained in automatic dictionaries for MT systems. The needs of translators are quite different. Translators rarely need to know about grammatical functions, syntactic and semantic categories, inflected forms, etc. – all data essential for automatic analysis processes. Translators can consult under 'base forms'; they do not need access from all variants. Nor do translators often need to consult dictionaries for items of general vocabulary. The primary need is to find the appropriate equivalents of specialist and technical terms. What is needed is reliable information about precise meanings, connotations, ranges of application, authoritative definitions and, if possible, examples of actual usage.

In the earlier days of MT these differences were not appreciated, and a number of researchers suggested that the MT dictionaries they were compiling could be used as machine aids for translators. Now it is recognised that the aims and methodology of automatic dictionaries and term banks has to be quite separate from MT research. This has been the reason why they have not been treated in this account of MT development (for references on term banks and similar machine aids see Bruderer 1978, Hutchins 1978, Snell 1979, Aslib 1981, Lawson 1982, Lawson 1985).[3]

17. 7: Interactive systems.

Automatic dictionaries and computer-based terminology databanks represent the lowest level of man-machine collaboration in translation; the human translator still undertakes the essence

[3] See also: J.Hutchins 'The origins of the translator's workstation', *Machine Translation* 13 (4), 1998, p.287-307.

of the translation process, the conversion of a SL text into a TL text. At the next level of man-machine collaboration, the computer provides versions of input sentences which the translator can improve.

The dividing line between human-aided (interactive) MT and machine-aided translation is ill-defined; in broad terms, we may agree that the crucial question is whether the computer translates individual lexical entries, which may include complex phrases as well as single words, or attempts to translate whole sentences. In the first case, human assistance involves the combination of lexical elements into coherent texts; in the second, human assistance involves the correction, improvement or editing of a computer-generated text. In the first case, the human translator is aided by a computer-based dictionary, i.e. it is 'machine-aided translation'. In the second case, the basic (even if rudimentary) translation process is undertaken by the computer, it is human-aided MT.

Interactive (human-aided) MT has been made possible by the development of on-line and text-editing facilities. There are now a number of systems on the commercial market, and no doubt there will be many more in the future. The range of collaboration is considerable; some systems demand intervention at many stages of the translation process, others only to help solve problems of ambiguity. The basic features of interactive MT, however, are that the system provides more than dictionary equivalents and that there is some interrogation initiated by the computer system itself.

17. 8: RAND Corporation, the MIND system.

In the late 1960s and early 1970s an important project in computational linguistics was the MIND (Management of Information through Natural Discourse) system at Rand Corporation. Its chief designers were Martin Kay, who had been a member of the Cambridge Language Research Unit (Ch.5.2), and Ronald Kaplan. MIND was intended as a general-purpose language data processing system comprising "an extensible set of fundamental linguistic processors that can be combined on command to carry out a great variety of tasks from grammar testing to question-answering and language translation." (Kay 1973a). Its components were: a morphological analyser, a syntactic processor, a disambiguator, a semantic processor, and an output component. The syntactic processor was an early and influential implementation of the 'chart' technique, devised by Kay (1973a) and Kaplan (1973) for parsing natural language (Ch.9.15). The semantic processor was intended as a means of representing semantic relationships (logical and thesaural) and of deriving semantic representations of input sentences.

The disambiguator was a system for dealing with problems of ambiguity not resolved by the morphological and syntactic components. It was envisaged as an interactive system, the computer asking users to assist in solving problems. For example, acting as a MT system (English-French) and encountering a sentence such as *They filled the tank with gas* the system might ask a human consultant:

DOES THE WORD 'TANK' REFER TO
1. A MILITARY VEHICLE?
2. A VESSEL FOR FLUIDS?

If the user types "1" the system would translate *tank* into French as *char d'assaut*, otherwise it would translate it as *tanque*. The system might then ask further questions:

DOES 'GAS' REFER TO
1. GASOLINE?
2. VAPOR?

(If the answer is "1" the translation would be *essence*, if "2" *gaz*.)

DOES 'THEY' REFER TO
1. THE SOLDIERS?
2. THE TANKS?
3. THE SHELLS?

4. THE ENEMY? (or any other recently used noun)

As well as such problems of homonymy and pronominal reference, the user might be asked to resolve syntactic ambiguities, e.g. in the case of *He saw the girl with the telescope*:

DOES THIS MEAN:
1. 'SAW WITH THE TELESCOPE'?
2. 'GIRL WITH THE TELESCOPE'?

The disambiguator itself decided what the problems were and what set of questions would resolve them in the most efficient way.

In general approach there were similarities with the earlier interactive CLRU experiment (Ch.5.2). Although the MIND project remained strictly a 'laboratory' experiment, its approach has been adopted by many later interactive systems.

17. 9: Chinese University of Hong Kong (1972-

Research on Chinese-English MT was begun by Shiu-Chang Loh at the Chinese University of Hong Kong in late 1969. A prototype system for translating Chinese mathematics texts into English was programmed on an ICL 1904A in October 1972. In the following years it developed into the CULT system (Chinese University Language Translator), which, in January 1975, with the sponsorship of the Asia Foundation (and later the Rockefeller Brothers Fund), began to be used regularly for the production of a cover-to-cover translation of the mathematical journal *Acta Mathematica Sinica* published in Peking. (For a brief period CULT was also used to translate physics articles from *Acta Physica Sinica*.)

CULT is an interactive on-line MT system, based on the 'direct translation' strategy (Loh & Kong 1977). Sentences are analysed and translated one at a time in a series of passes, and the parser identifies little more than 'surface' groupings of grammatical categories (noun and verb groups, modifiers). After each pass a portion of the sentence is translated into English, the human 'editor' intervening for the insertion of articles, the determination of voice and tense of verbs, and the resolution of semantic and syntactic ambiguities. If the parser fails, the 'editor' adds entries to the dictionary and may translate the whole sentence. Although analysis and synthesis are integrated there are, however, identifiable stages of the translation process (Loh & Kong 1979, Loh et al. 1979): source text preparation; input (via Chinese keyboard); lexical analysis (i.e. SL dictionary lookup, including on-line updating of dictionary); syntactic and semantic analysis (determination of grammatical categories, identification of noun groups, verb groups, modifier groups and auxiliaries); relative order analysis (recognition of relationships among nouns groups, verb groups and modifiers); target equivalence analysis (resolution of multiple meanings); output (TL dictionary lookup); output refinement (reformatting, insertion of mathematical formulae).

CULT was originally implemented in batch mode, where a major problem was that of inputting Chinese characters. They had to be manually encoded into four-bit standard telegraph code and punched onto cards. The process was both laborious and prone to considerable error. For this reason, Loh devoted considerable effort to the development of a Chinese keyboard, which enabled on-line input and a genuine interactive system. CULT is now implemented on a PDP11/34.

Development in subsequent years has apparently concentrated mainly on dictionaries, initially from the need to compile up-to-date glossaries of Chinese mathematical terms. Later, there has been some experimental work on an extension of CULT to English-Chinese translation, and this has involved the reorganisation of dictionary structures to facilitate bi-directional translation (Loh & Kong 1979, Loh et al. 1984).

As an interactive MT system CULT was a pioneer, conceived long before on-line systems were common. It was also designed to translate specific types of texts and not as a general purpose system. In consequence CULT appears somewhat crude in comparison with later interactive MT systems such as ALPS and Weidner (below). The designers of CULT had no pretensions about achieving high-quality MT; they were interested simply in producing 'translations' for a specific practical purpose. In this respect CULT is successful: the system works, translations are produced, and the publication is sold. An example from *Acta Mathematica Sinica* vol.18, 1975:

> IT IS GENERALLY KNOWN THAT, THE MAXIMAL CONDITION AND THE MINIMAL CONDITION FOR RINGS ARE MUTUALLY INDEPENDENT, BUT IN MANY PROPERTIES OF RINGS THERE ARE OFTEN CASES IN WHICH THEY ARE MUTUALLY RELATED. FOR EXAMPLE ASOCIATIVE RINGS SATISFYING THE MAXIMAL CONDITION OR THE MINIMAL CONDITION, THEIR NIL RADICALS ARE NILPOTENT. PARTICULARLY IN 1939 C.HOPKING POINTS OUT, SO LONG AS AN ASSOCIATIVE RING CONTAINS A LEFT UNIT ELEMENT, THEN THE MINIMAL CONDITION FOR ITS LEFT IDEALS MUST IMPLY THE MAXIMAL CONDITION FOR LEFT IDEALS. UP TO NOW THIS IS THE MOST GENERAL RESULT, IN A COMMUTATIVE ASOCIATIVE RING CONTAINING A UNIT ELEMENT COHEN HAS SIMILAR RESULTS. THESE RESULTS SUFFICIENTLY ILUSTRATE THAT BETWEEN THE MINIMAL CONDITION AND THE MAXIMAL CONDITION THERE IS AN ESSENTIAL ASSOCIATION AND THEIR MUTUAL EXCHANGE RELATION. DIALECTICAL MATERIALISM IS IN CONTRAST TO METAPHYSICS. IT REGARDS THAT ALL MATTERS ARE IN MUTUAL ASSOCIATION AND MUTUAL CONSTRAINT. CONTRADICTION AND OPPOSITION CAN ALL MUTUALLY BE EXCHANGED UNDER DEFINITE CONDITIONS. THE PROBLEM IS HOW TO GRASP DIALECTICALLY THE CONDITION WHICH MAKES OPPOSITION EXCHANGE MUTUALLY....

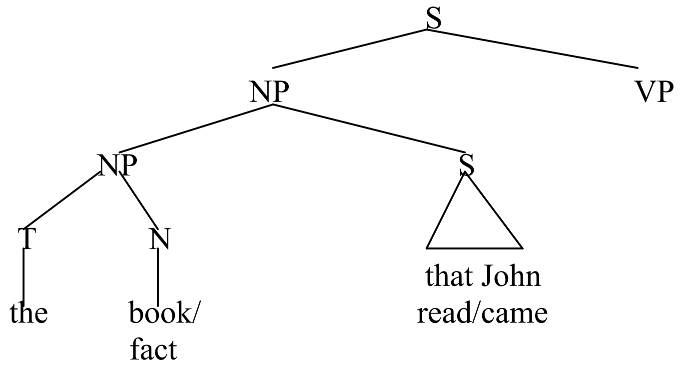
17. 10: Brigham Young University (1970-79)

The Translation Sciences Institute was established in 1970 under the leadership of Eldon G.Lytle at Brigham Young University (BYU), Provo, Utah, funded by the University and the Church of Jesus Christ of Latter-day Saints (the Mormons). The project was the development of an interactive MT system for translating English texts into a number of languages (initially French, Spanish, Portuguese, German and Chinese), to be called the Interactive Translation System (ITS). The system was based on Lytle's theory of 'junction grammar', a modification (in parts radical) of transformational grammar, which was to provide the formalism for 'transfer' representations of SL sentences (Lytle 1974, Lytle et al.1975, Lytle 1980, Melby 1980a, 1980b). English analysis was to be performed interactively at a terminal, linked to an IBM mainframe computer; multilingual TL synthesis was to be fully automatic. It was argued that accurate unambiguous 'transfer' representations could be produced only by making use of the linguistic knowledge of human operators, and that the amount of interaction involved could be justified only if multiple TL texts were output.

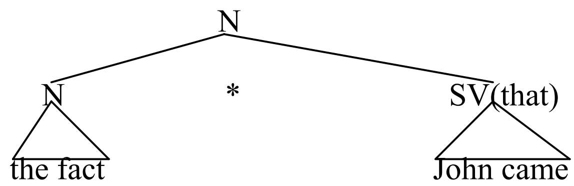
The BYU system ITS had the usual stages of a 'transfer' system. Analysis had three phases: dictionary lookup; division of sentences into 'phrases' by making cuts after each noun, e.g. *(The boy) (in the car) (ate a very good hamburger) (while driving down the street)* – cf. Wilks' similar procedure (Ch.15.1 above); and syntactic analysis to produce junction grammar tree-representations, during which both lexical ambiguities (e.g. homographs) and structural ambiguities, (e.g. of prepositional phrases) were resolved in interaction with human operators. Thus, given *The boy threw the ball on the table*, the operator would be asked, as in the MIND system, 'What does the prepositional object modify?').

The Transfer algorithm converted SL junction trees into TL junction trees, adjusting as necessary by reference to TL dictionary information. For example English V$(P+N), as for *to stare at X*, became French (V+N)$(A+E), *regarder X fixement*. Finally, Synthesis included lexical reordering rules and morphological formation.

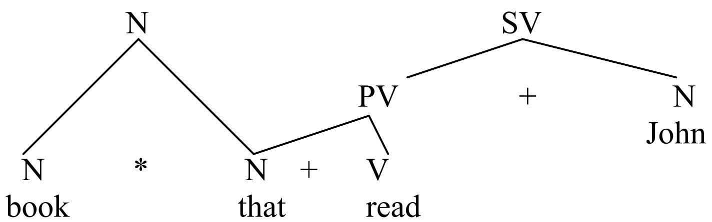
Most research activity was devoted to the development of junction grammar and the interactive analysis routines. The distinctive feature of Lytle's theory was the treatment of relative clauses (Lytle et al. 1975). In transformational grammar (cf. Ch.3.5 above) they are regarded as subordinate constituents of noun phrases whatever the semantic relationship, e.g. both *The book that John read...* and *The fact that John came...* would be represented as:



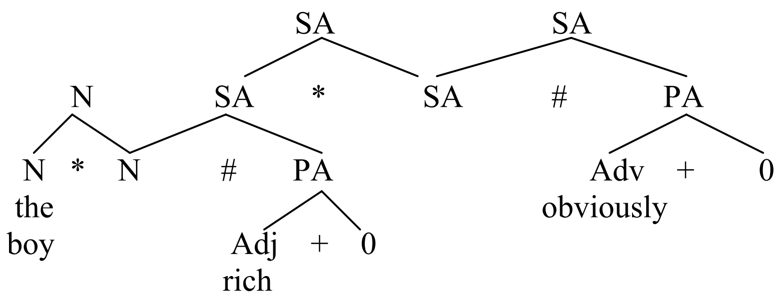
Junction grammar distinguishes between 'subjunction' (indicated by *) as in *the fact that John came:*



and 'interjunction' (indicated by +) as in *the book that John read*:



In this way, the two interpretations of *The fact that he learned surprised us* can be distinguished. Interjunctions were also identified within noun phrases, e.g. the obviously rich boy (Lytle 1980 p.330):



In interactive analysis, the operator would be asked to specify what type of relationship pertains to a given structure - i.e. knowledge of junction grammar was a prerequisite for ITS analysts. Given the complexity of the interaction it was obviously essential, on cost grounds if for no other, to work towards multilingual synthesis.

An evaluation of ITS in 1979 by its sponsors concluded that the system had little commercial possibilities. The system had become too elaborate, it was decided, and it was more practical to develop translator aids. With other members of the project Lytle left BYU for the private sector and established ALPS (below). The company which had been set up to market the BYU system, Computer Translation Inc. (CTI), found it had no product and turned successfully to selling microcomputers (Beesley 1982).

After the formation of ALPS, Alan K.Melby continued working at Brigham Young on the development of a new version of ITS (Melby 1982, 1983, 1984). Translators were feeling as if they were working for the machine; ITS forced them to answer "many uninteresting questions and to revise many sentences they thought should be retranslated"; in the new version, the translators are to be in control, making the machine work for them. Melby proposes a three-level approach: at the first level translators operate a word processor with optional access to an automatic bilingual dictionary or (multilingual) terminology databank; level two adds dynamic processing of SL texts, the automatic searching of term files and display of possible TL equivalents, and the automatic amendment and on-line editing of TL texts; level three integrates a full MT system to give translations which may be accepted, revised or rejected as appropriate (in some cases the MT system may fail completely, in others it may have only partial success). The aim is to please both users "because they are in control and choose the level of aid" and designers of MT systems "because they are not pressured to make compromises just to get automatic translation on every sentence". The system integrates a MT system and a terminology aid system, representing a shift of direction which was also followed by those BYU members who went to ALPS and by the rival group in Provo, Weidner Communications (below). Melby himself has not researched the MT system but has concentrated on the design of the workstation.[4]

17. 11: Automated Language Processing System, ALPS (1980-

ALPS was formed in January 1980 by members of the Brigham Young University group (including Eldon Lytle) in order to continue commercial development of their interactive approach to MT. Although its headquarters are in Provo, Utah, the company has no financial ties with either the University or the Mormon Church. The ALPS system came on to the market in 1983 and in 1984 the company opened a European office in Neuchâtel, Switzerland.[5]

The ALPS system is designed to provide translators with a wide range of computer aids with good text editing and dictionary facilities (Danik 1984, Kingscott 1984). Developed for Data General equipment, it is intended in time to be 'hardware independent'. The basic component is a multilingual word processor (i.e. one capable of handling all Roman scripts), linked to a printer. The user can choose to augment this facility by

(i) Selective Dictionary Lookup, which when required accesses the users' own dictionaries, which may be monolingual, bilingual or multilingual according to preference, and displays the entries.
(ii) Automatic Dictionary Lookup, in which the computer searches dictionaries for SL words and displays TL equivalents (in 1984 available for English, French, German, Spanish and Italian). Words found are listed with contexts, words not found are listed for dictionary updating. No

[4] See also: J.Hutchins 'The origins of the translator's workstation', *Machine Translation* 13 (4), 1998, p.287-307.
[5] Shortly afterwards, Automated Language Processing Systems withdrew from sales of its system. It became primarily a translation service, was renamed ALPNET, and acquired other translation companies. Subsequently, it used its MT systems internally only. See A.T.Zirkle 'The role of computer-aided translation in translation services', *Translating and the computer 10: the translation environment 10 years on. Proceedings of a conference... 10-11 November 1988, CBI Conference Centre, London*, ed. P.Mayorcas (London: Aslib, 1990), 11-17.

syntactic processing takes place, but there is automatic morphological analysis to determine stem forms.
(iii) Computer Translation System (CTS), which creates a working dictionary of SL text words, asks translators about ambiguities of spelling, lexical choices and syntactic problems (as in TSI, above). For example, it would display the possible TL equivalents for a SL word and ask translators to select one, or if the word is not in the dictionary to provide one. (In 1984 the CTS was available for the pairs English-French, English-German, English-Spanish, English-Italian, and vice versa.)

The emphasis is on close involvement of translators, a user-friendly, interactive system, which users can adapt to specific needs (from simple dictionary lookup to unedited translation) and to organisational requirements, e.g. by developing personal programs (particularly through the word processing software) and interfacing with other computer equipment such as OCR readers (Danik 1984).[6]

In an effort to simplify the translators' tasks and to retain their confidence that they are "not merely... proofreading the computer's work, but.. actively participating in the production" (that "the translator is in control"), the ALPS system has reduced the complexity of dictionary information to a minimum. The argument is that technical linguistic knowledge cannot be assumed, that dictionary compilation should not be onerous, and that any information given by translators should be seen to be actually used. There is no need to incorporate sophisticated semantics and AI-type knowledge representations in the dictionaries; the translators themselves are the best sources of this kind of information (Lonsdale 1984). As a consequence, the dictionary information appears to be minimal, little more than data on morphological variants, grammatical categories and potential TL equivalents. The system comes with a starter dictionary of 10,000 base forms, representing (it is claimed) over 90% of the running words of most general texts (Wood & Wright 1985).

It is not only dictionary information which has been kept simple. It would appear that syntactic processing has also been simplified. It is argued that questions during SL analysis should be as succinct and unambiguous as possible, and that translators should not be asked 'trivial' questions of syntax (Lonsdale 1984). By reducing the complexity of analysis programs, the ALPS system has less need for powerful processing algorithms. Although details of ALPS syntactic processing are not available (for god commercial reasons), it certainly includes "inflection, conjugation, agreement, and word ordering" of some unspecified kind (Danik 1984). However, it would appear that the BYU junction grammar plays no longer a significant role (Beesley 1982)

With its emphasis on interaction, simplified grammatical data in dictionaries and rudimentary syntactic analysis, ALPS would appear to be not a great deal more than a sophisticated text processor with immediate access to bilingual dictionaries. No doubt good results can be achieved, and translators' time can be saved, but it is almost impossible to assess the relative contributions of the system itself and of the translators to the actual translation processes and to the final products.

17. 12: Weidner Communications Corporation (1977-

The Weidner Communications Corporation (WCC) was established in Provo, Utah, and for some time was in competition with the Translation Sciences Institute of Brigham Young University. The rivalry, occasionally bitter it seems, has continued to exist between Weidner and the ALPS group (Beesley 1982). Although founded by members of the Mormon Church, Weidner is a private company with no financial links with either Brigham Young University or the Church. In 1982 WCC moved its headquarters to Northbrook, Illinois, near Chicago, while retaining its Research and Development Center in Provo. In 1984 Weidner was acquired by Bravis International, one of Japan's largest translation companies.[7]

---

[6] Subsequently, the ALPSystems incorporated an early version of a translation memory (a 'repetitions file') – see J.Hutchins 'The origins of the translator's workstation', *Machine Translation* 13 (4), 1998, p.287-307.
[7] The company ceased to market the Weidner systems in 1990.

From the beginning the aim was a system for computer aided translation (CAT), a text processing system linked to an interactive MT system, with dictionary updating by translators. In brief, the translator (or typist) enters the SL text at a terminal, words are checked in the dictionary, a translation appears sentence by sentence on the screen, corrections are made to the text using the word processing software, and the corrected TL text is printed out.[8]

The WCC translation system itself operates in the following steps (*LM* 10, July 1984, p.12-13; WCC handout 1984). Exact details are not available for obvious commercial reasons, and the precise sequence of some stages is uncertain. The 'Input Handler' controls SL text input and breaks up text into words and sentences; 'Dictionary lookup' performs morphological analysis (removing endings) and searches the dictionary; and the 'Idiom Converter' recognises idiomatic expressions entered by translators in the dictionary. Then follow stages of 'Homograph resolution' (consulting contexts as indicated in the dictionary), and 'Syntactic analysis' for the identification of noun and verb phrases and recognition of interrelationships in order to form some kind of phrase structure analysis. The final stages of transfer and synthesis includes routines for reordering text according to TL structures, 'Verb conjugation' and 'Agreement' (i.e. morphological synthesis of TL verb and noun forms), 'Cosmetic routines' for inserting TL articles and prepositions as necessary (according to dictionary information). Finally, an 'Output Handler' reorganises the text into a format corresponding to the SL original.

The TL text is then passed to the 'Text Amender', a word processor system for editing. To assist revision the translator has access to a 'Synonym dictionary', alternative TL expressions which have been entered during dictionary compilation. Any items of vocabulary not found during the translation process can be displayed for the translator either alphabetically, by frequency in the text, in text order or by context. To save translators time, dictionary searching can be done as a separate operation, e.g. overnight.

As in all MT systems, dictionary quality is vital for good results. The WCC dictionary compilation program ensures consistency of dictionary information by asking translators for specific data about the vocabulary being entered. Hundt (1982) gives some details. The information required for nouns includes a semantic category (e.g. Human Group, Body Part, Animal, Inanimate, Concrete, Abstract) as well as gender and number, whether it is a proper noun or a noun indicating time or place, and its inflection (the compiler is asked to select the relevant paradigm). For verbs the information includes data on its conjugation (selected from paradigms, 42 in the case of German), valency, transitivity, reflexivity, type of direct object, whether it takes an indirect object, whether its prefix is separable, whether its present and past participle forms can be adjectives. In addition, dictionary entries include information on TL word order, e.g. in the case of an adjective whether it follows nouns in TL. In the case of idiomatic expressions (which of course may include specialised terminology as well as true idioms), entries have to be made under each of the component items and indicate which of them are invariant in SL and in TL. Where a particular SL word or expression can have more than one TL equivalent in different contexts, the possible alternatives (up to six) are entered in the Synonym dictionary which is accessed during the TL text editing processes.

Weidner market the system on a 'turnkey' basis at a fixed price for a language pair. There are now two versions: MacroCAT running on a DEC PDP11 or DEC Vax computer, and MicroCAT running on an IBM PC microcomputer. Both packages include computers, terminals, the translation software, the text processing software, a printer, and a dictionary of current vocabulary containing 8000 stems and 3000 expressions. MacroCAT translates at a rate of 14,000 words per hour, and MicroCAT at 1,600 an hour. In 1985 the language pairs available were English-Spanish, Spanish-English, English-French, English-Arabic (not MicroCAT), English-

[8] The categorisation of the Weidner systems as 'interactive' was inaccurate. They were batch-processing systems of the 'direct translation' type with facilities for post-editing and dictionary updating. The fact that texts could be processed sentence by sentence did not make them truly 'interactve'.

German, English-Portuguese, English-Italian, with versions under development for French-English, German-English, German-French and French-Spanish, and (since the acquisition by Bravis) for Japanese-English.[9] For both versions there is available a maintenance agreement which provides for the delivery of improved versions of the software every six months.

The WCC system is not claimed to be more than a 'computer-aided translation' system and there is no expectation of good quality output without editing. Nevertheless, 'raw' translations do not appear to present particular difficulties to translators when revising. In a 1980 evaluation Pigott considered that the "quality of the Weidner Spanish-English system was comparable to that of Systran French-English" but the quality of the English-French and English-Spanish versions was "markedly poorer than that of the Systran English-French and English-Italian" (Van Slype 1983) Nevertheless, post-dating time was a third of the time required for full human translation of the same text, and so there were considerable benefits. WCC claim that revisers can work at a rate of 600 to 1000 words an hour, which users such as the Mitel Corporation in Canada can corroborate (Hundt 1982).

Some example unedited translations into English are the following (from ITTE Technical Translations & Linguistics Centre, Feb 1984):

French-English:

> This function assigns ports to the exchanges and to the workstations by the parameters introduction. She allows also to put these circuits into service or out of service and authorises the ports of the workstations to carry out some functions or all functions. At last she allows to assign periodically recording equipment for the output of the message.

Spanish-English:

> Because of it these functions were considered as of prime importance during the development of the digital exchange ITT 1240. Taking advantage of to the maximum the inherent advantages of the digital technology and of the distributed control, the majority of the facilities of operation and maintenance has been incorporated directly to the equipment.

German-English:

> The automatic switching on of the operating voltage takes place after connection the supply voltage of the exchange after short voltage failures, after voltage fades and after the failure of a supply voltage branch. The manual switching on of a supply voltage branch is caused by pressing of the key "voltage converter" at the signal field (MAP).

17. 13: Other interactive systems

A number of projects in Japan are developing interactive MT systems (Ch.18.2ff below). It is interesting to note three recently established projects outside Japan are investigating English-Japanese interactive systems.

One project was established in 1985 at the Centre for Computational Linguistics in the University of Manchester Institute of Science and Technology (a centre already involved in the Eurotra project, Ch.14.2 above), with the sponsorship of the British Information Technology Programme (Alvey) and International Computers Ltd. The system design envisages a 'transfer' model where the human expert assists during the analysis and transfer stages by providing information derived from his SL knowledge and his knowledge of the subject matter (Johnson & Whitelock 1985). In this way, it is believed that a writer knowing only English should be able to produce high quality Japanese text.

The project is related to the proposed implementation of a Japanese-English analyser at the Department of Japanese Studies of Sheffield University (Jelinek and Hawgood 1984). The analyser

[9] When Bravis ceased marketing the systems, MacroCAT was acquired by the Intergraph Corporation and sold as DP/Translator, and later renamed Transcend.

was developed and programmed by Jiri Jelinek at Prague University before 1968. At Sheffield it has been successfully adapted in a printed form (the Integrated Japanese-English Grammar Dictionary) for teaching English businessmen and scientists to read Japanese. It is expected that the experience with learners and in particular with their ability to cope with multiple choices (in grammar and vocabulary) will be of assistance in the design of an interactive system based on the same principles.

At the Carnegie-Mellon University, Pittsburgh, Jaime Carbonell has recently revived his knowledge-based MT project (Ch.15.2 above) which he investigated at Yale, although now as the foundation for operation in an interactive mode. It is planned to integrate the approach with the interactive system of Masuru Tomita which was developed with Nishida and Doshita at Kyoto University (Ch.18.6 below). The aim is an English-Japanese system for composing and simultaneously translating short personal texts in interactive mode (Carbonell & Tomita 1985; Tomita 1985)