[Published in: *Machine Translation,* vol.13, no.4 (1998), p. 287-307]

The origins of the translator's workstation

John Hutchins
(University of East Anglia, Norwich NR4 7TJ, U.K.)
[E-mail: WJHutchins@compuserve.com]

Abstract

The first proposals for various component tools of what is now called the 'translator's workstation' or 'workbench' are traced back to the 1970s and early 1980s in various, often independent, proposals at different stages in the development of computers and in their use by translators.

Keywords: translation tools, workstation, translation memory, history

1. Introduction

The rapid and high profile emergence on the market in the 1990s of the workstation for the professional translator has given rise to certain assumptions by many of its users and advocates (e.g. Gordon, 1997). The first is that the concept of the 'translator's workstation' (or 'workbench'), and in particular its most distinctive component, the 'translation memory' (TM), are creations of the last ten years or so. The second is that the workstation and the translation memory spell the end of traditional machine translation (MT). The falsity of the second assumption can be demonstrated by, on the one hand, the increasing use of MT for technical documentation in multinational organisations (e.g. Brace et al., 1995) and by, on the other hand, the even greater increase in the use of unedited MT (or 'crummy' MT – Church and Hovy, 1993), particularly on the Internet for electronic mail, discussion groups, etc. (Flanagan, 1996), for translating Web pages, for accessing databases in foreign languages, and so forth.

As for the first assumption – that the translator's workstation is something new in the 1990s – it will be seen in this paper that proposals for virtually all components (including the translation memory) can be traced back over more than 20 years. Their full integration and acceptance had to await technical developments of the last decade, but their desirability for the effective utilisation of machine aids and translation tools was recognised long ago.

The title of 'workstation' has been applied to a number of translation aids that will not be the subject of this paper. Here we are concerned only with the type of workstation intended for direct use by professional translators knowing both source and target languages, and retaining full control over the production of their translations. Workstations and other computer-based translation tools are traditionally referred to as systems for 'machine aided human translation' (MAHT), in order to distinguish them from MT systems with some kind of human assistance either before or after processing (pre- and post-editing), known often as 'human-aided machine translation' (HAMT).

Other types of 'workstation' will not be treated. Firstly, there is the workbench built around fully automatic MT software which provides facilities for operators to prepare input texts, to post-edit output and to produce quality publications. This has been particularly popular in Japan where the producers of MT software have also been manufacturers of computer hardware, e.g. the 'workbench' from Toshiba (Amano et al., 1989) and the 'Engineering Workstation' from NEC (also from the late 1980s) described as a translation support system with facilities for dictionary generation and editing, textual pre-editing, and textual post-editing and capable of interfacing with OCR and DTP systems. In such configurations, the operator – who may or may not be a qualified translator – acts in effect as an 'assistant' to the machine, i.e. HAMT. The operator is not necessarily offered other translation tools such as terminology management and translation memory programs. (During the 1990s most of these systems have been sold for personal computers.)

Another type of 'workstation', investigated by a number of researchers (e.g. Boitet, 1990; Huang, 1990; Jacqmin, 1992; Jones and Tsujii, 1990; Somers et al., 1990) but also not treated here, would enable a monolingual user to compose documents (such as letters and short reports) in a language known only poorly or not at all – or even, in more than one unknown language. The user would interact closely with the computer to create a text – perhaps using standard templates – that the translation software would be able to deal with at a predictably high quality level. In most cases, the subject range would be restricted (and perhaps also the stylistic range) in order to minimize problems of ambiguity in the source language and problems of lexical selection in the target language (or languages) and in order to ensure generation of idiomatic sentences. In such systems there may not be a 'finished' initial source text; instead target texts are composed by human-computer 'dialogue' at the workstation.

Finally, we will not be treating here the idea of a 'workstation' for the use of researchers who are developing MT software (Nirenburg, 1992). In this case, the facilities must include sophisticated linguistic resources, support for knowledge acquisition, e.g. for dictionary and grammar creation, and appropriate means for presenting information about any aspect of the translation processes under development. Similarly, we are not concerned with 'workstations' for the acquisition, development and maintenance of lexicographical, lexicological and terminological resources, or for research on text corpora, whether monolingual, bilingual or multilingual, and for whatever purpose.

In this survey of origins, we are concerned therefore only with components which have been integrated in a range of now commercially available workstations intended for the professional translator (MAHT). These workstations make available to the translator at one terminal (whether individual computer or as part of a company network) a range of integrated facilities: multilingual word processing, OCR scanning, electronic transmission and receipt of documents, spelling and grammar checkers (and perhaps style checkers or drafting aids), publication software, terminology management, text concordancing software, access to local or remote termbanks (or other resources), translation memory (for access to individual or corporate translations), and access to automatic translation software. Initially, not all commercial workstations offered all these facilities, but in the course of the early 1990s most introduced the full range.

The paper focuses on the origins of the fundamental concepts underlying the basic components of the professional translator's workstation. It does not attempt to describe such developments as when particular translation tools were developed and available commercially, whether they were successful (in practice or on the marketplace), how they were integrated with other tools, what the first translator's workstations could or could not accomplish, how they were received by the profession, etc. As will be seen, different people from different backgrounds came to similar and usually independent proposals at different times and at different stages of the development of computers. While it is not possible to point to any single conceptual origin for any one feature, it can at least be demonstrated that the basic concepts antedate by many years the commercialisation of translator's workstations in the last five years.

2. ALPAC

In 1966, research on machine translation received a major setback with the publication of a report sponsored by the major US funding agencies. They had asked the National Academy of Sciences' National Research Council to investigate the prospects of operational MT systems in the near future. The report of the Automatic Language Processing Advisory Committee (ALPAC, 1966) was a blow to those researchers who were aiming for general-purpose fully automatic translation of unrestricted text. The consequences of the report were massive reductions in the support for MT in the United States.

However, ALPAC was not entirely negative; it encouraged support for basic computational linguistics and, in addition (today often forgotten), the development of computer-based aids for translators. Members of the committee had been particularly impressed by the aids provided at the German Federal Armed Forces Translation Agency (Bundessprachenamt) in Mannheim. Here, computers were being used to produce text-related glossaries. These were lists of the technical words appearing in the source texts on which translators were currently working, together with their equivalents in the target language. The committee was also impressed by the use of computers at the European Coal and Steel Community to produce a terminological database for translators, which included terms in their sentence contexts (this was DICAUTOM, a precursor of the Eurodicautom of the European Commission.) It believed that aids such as these could be much more economically effective in the support of translation production than any of the current MT systems.

This view was supported by translators themselves. They were generally antagonistic to the very idea of machine translation. Not only was it seen as a direct threat to their livelihood, but translators were being employed to improve the poor and sometimes unreadable output from MT systems. This situation continued into the 1970s and later. Automatic translation was definitely not seen as any help for the professional translator at all. At this time, the only examples of working MT systems were 'batch' systems. The output was in the form of unfriendly computer printout on lined paper – often all letters were capitalised, and much of the output was almost illegible. Post-editing meant the correction by hand (pencil or biro) of the raw text, and the handing over of marked-up sheets to a typist. The work involved was quite different from the kind of revision that was the norm for human translations. Revisers were not improving style and correcting inconsistencies of terminology, but in effect doing 're-translations'. The mistakes made by the MT systems were far more basic

than those that even the worst human translator would commit: gross grammatical errors, distorted syntax, inappropriate lexical choices, and more. In effect, the revisers were being asked to be 'slaves' to the machine – and, quite understandably, they did not like it.

3. Term banks and text-related glossaries

Computer-based terminological resources were, by contrast, received with increasing favour by translators from the late 1960s. Particularly in large governmental and industrial organisations, there was an increasingly pressing need for fast access to up-to-date glossaries and dictionaries in science, technology, economics and the social sciences in general. The difficulties were clear: rapidly changing terminology in many scientific and technical disciplines, the emergence of new concepts, new techniques and new products, the often insufficient standardisation of terminology, and the multiplicity of information sources of variable quality and reliability. It had been estimated that translators might be spending up to 60% of their time consulting dictionaries, glossaries and other terminological sources (Krollmann et al., 1965)

It was recognised from the outset that on-line dictionaries for translators could not be the kinds of dictionaries developed in MT systems—as Oettinger had believed in the 1950s (Oettinger, 1960). Translators do not need the kind of detailed information about grammatical functions, syntactic categories, semantic features, inflected forms, etc. which is to be found in MT lexica, and which is indeed essential for automatic analysis. Nor do translators need to consult dictionaries for items of general vocabulary–which are equally essential components of a MT system dealing with full sentences. The primary need is for access to specialised technical and scientific terminology with translations into standard 'approved' equivalents.

In the 1970s, terminology data banks were being built to provide information on demand about individual words or phrases (definitions, examples, translations), as the basis for the production of glossaries for specific texts, and for the production of published up-to-date specialised dictionaries for general use (Hutchins, 1978: 142-146). Many of the databanks were multilingual, nearly all provided direct on-line access and most included definitions, and some were very large, e.g. the multilingual TEAM database at Siemens contained some 700,000 entries covering more than 2 million scientific and technical terms in the major European languages.

In the case of other termbanks, the emphasis was on the provision of terms in actual context. For example, in Eurodicautom (developed for the translation service of the European Commission), new terms were supplied with illustrative passages, texts in another language with an equivalent expression, definitions from reliable sources, subject field codes, and bibliographical references. Likewise, contexts and authoritative definitions were provided for the English and French terms in the Canadian TERMIUM databank, established in 1970 at the University of Montreal as a central repository of terminology for the country's translation services. Both systems were accessible on-line at this time.

4. Friedrich Krollmann

At the German Army Translation Service (later Bundessprachenamt), the LEXIS system had been developed under Friedrich Krollmann since 1965 for both the production of up-to-date printed dictionaries and word lists of scientific and technical

terms and the provision of text-related glossaries. The latter were produced on demand for translators working on particular texts. Translators would mark in their source texts those words they needed help with. The computer provided a printout of the original terms and any translations found in the databank, either in the order in which they occurred in the original source or in alphabetical order. The facility was found particularly useful when a team of translators was working on a single large project. The experience of the Bundessprachenamt over many years was that "text-related glossaries can be a decisive aid towards improving translation work both quantitatively and qualitatively" (Krollmann, 1974: 127) – one of the earliest confirmations of the considerable economic advantages of appropriate computer facilities in day-to-day translation work.

However, Krollmann (1971) envisaged further expansion of the terminological database into a 'linguistic data bank' to support translators.

> The model of a linguistic data bank for the translator consists of eight branches or subbanks, one of them being exclusively a program bank from which various subprograms… or routines may be called upon to operate the other data-subbanks or run linguistic processes according to specific needs.
>
> The other subbanks are divided as follows:
>
> I. Multilingual dictionaries for lexicographical work;
>
> II. Multilingual dictionaries for machine-aided translation;
>
> III. Monolingual dictionaries whose entries contain additional information on Subbank I entries, particularly definitions (so-called background storage);
>
> IV. Thesauri, particularly for information retrieval and documentation purposes;
>
> V. Index file for registering and checking all translations in order to avoid duplication;
>
> VI. Corpus bank for analyzing texts according to linguistic criteria;
>
> VII. Translation archives.
>
> (Krollmann 1971: 118f)

The databases were intended not just for translators but also for lexicographers and other documentation workers, with facilities for compiling dictionaries and term glossaries, for producing text-related glossaries for machine-aided translation, for direct on-line access to multilingual terminology databanks, and for accessing already translated texts by means of indexes. The archive of translations, recorded on magnetic tapes, could also be the source of re-usable translation segments. Krollmann envisaged that:

> … via descriptors or keywords, large batches of text could automatically be searched for particular passages and then be displayed on video screens as an aid to the translator; [and] for revised new editions of translations only the changed passages would have to be retyped. Insertion of changes and corrections into the old text would automatically be done by computer…
>
> (*ibid*., 123)

However, the whole complex of interlinked linguistic databases was constrained by the computer technology then available; and Krollmann stressed that the high costs demanded heavy usage – when "turnover is so large that it can no longer be handled manually", when "the use of a data bank is not only economical but has become a necessity." (*ibid*., 124)

5. Erhard Lippmann

There were already by this time developments that would revolutionise the use of computers. Krollmann had already seen the importance of ‘time-sharing’. It allowed a number of people to each communicate directly with a mainframe computer simultaneously from terminals, which could be at locations other than the building housing the computer. Access was provided by links over a telephone network. Now, users of computers did not any longer have to take their programs and data physically (e.g. on punched cards) to a computer centre, wait for a batch run to be submitted and completed, return at a later time (sometimes days later) to collect the results, check for mistakes, make changes and then resubmit again. Now the whole operation could be done from users’ own terminals; programs could be submitted directly, the results printed or displayed locally; users could call upon the resources of the computer almost on demand.

At the IBM research centre at Yorktown Heights, Erhard O. Lippmann investigated the possibilities of time-sharing for computer-aided translation. An internal report appeared in August 1969; it was published more widely in brief form in October 1970 (Lippmann and Plath, 1970) and then in full detail in February 1971 (Lippmann, 1971.) Time-sharing made possible the provision of:

> computer aids to enhance and accelerate human translation... by employing man-machine interaction techniques. In contrast to MT, such a system does not attempt to simulate the human translator by producing an autonomous translation... Instead, the system serves as an extension of the capabilities of the user. (Lippmann 1971: 10)

Lippmann summarised as follows:

> Rapid iteration toward the desired goal (i.e. a finished translation) can be achieved by switching back and forth as many times as required among human translation, direct dictionary lookup, editing, file management, and printing via typewriters or display screens. (*ibid*., 10f)

In addition, access could also be made to computer resources outside the organisation, e.g. to remote terminological databases.

Lippmann described how the translator of the future might work. Seated at a terminal, the translator would

> (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... (*ibid.*, 11)

However, many features were still undeveloped. Cursor control was by typed commands. It was not possible to consult the dictionary with the text still on screen: users had to save the file, load the dictionary file, search it, make a note of the translation required (on paper), exit the dictionary and then reload the text file. The screen did not display what the printer would produce. Printing commands had to be inserted in the text itself, diacritics were entered as character strings (ä appeared as 'a@'"), and an underlining such as *muss* would be entered 'muss@@@@____'. The deletion and substitution of words and phrases involved the retrieval of lines of text, typing the characters to be removed and typing their replacements. Equally laborious was the movement of segments of texts from one area of the file to another (there were no 'cut' and 'paste' facilities.)

Lippmann himself proposed facilities for improving on-line access to terminology by the generation of ‘terminological digests’ or “lists of terms extracted from a main dictionary in the order of textual occurrence or alphabetical occurrence”

(Lippmann, 1975: 310). Whereas, the text-related glossaries of the Bundessprachenamt were printed out, Lippmann described how they could be generated and displayed on the text screen and how a translator could "flip back and forth between his input and the… terminological digest terms." (*ibid*., 314)

Before Lippmann's ideas could be realised there needed to be great improvements in text processing. They were to come from the late 1970s onwards. Mechanisms for text processing at an individual terminal were very much at an early stage–much of Lippmann's paper is concerned with operations (deleting, moving, replacing and searching for words) that were to become commonplace for microcomputers within a decade.

6. Peter Arthern

The use of a translation archive was elaborated by Peter Arthern (1979) in a proposal for what has now, since the late 1980s, become known as a 'translation memory.' The suggestion was made in a discussion of the potential use of computer-based terminology systems in the European Commission. Arthern's paper argues forcefully that the use of unrevised MT in the Commission's translation services was premature—based on the evaluations of Systran, which had been performed at this date [1978]—although he did think there was scope for post-edited MT of a restricted range of texts (e.g. minutes of meetings) as long as the costs were low enough. After stressing the importance of developing multilingual text processing tools and of providing access to terminological databanks, he went on to comment that many European Commission texts were "highly repetitive, frequently quoting whole passages from existing Community documents" and that translators were wasting much time "re-translating texts which have been already translated" (Arthern 1979: 94). He proposed the storage of all source and translated texts, the ability to quickly retrieve any parts of any texts, and their immediate insertion into new documents as required. He referred to his concept as "translation by text-retrieval":

> The pre-requisite for implementing my proposal is that the text-processing system should have a large enough central memory store. If this is available, the proposal is simply that the organization in question should store all the texts it produces in the system's memory, together with their translations into however many languages are required.
>
> This information would have to be stored in such a way that any given portion of text in any of the languages involved can be located immediately… together with its translation… (Arthern 1979: 94f)

Arthern envisaged a workstation where

> simply by entering the final version of a text for printing, as prepared on the screen at the keyboard terminal, and indicating in which languages translations were required, the system would be instructed to compare the new text, probably sentence by sentence, with all the previously recorded texts prepared in the organization in that language, and to print out the nearest available equivalent for each sentence in all the target languages at the same time, on different printers...
>
> Depending on how much of the new original was already in store, the subsequent work on the target language texts would range from the insertion of names and dates in standard letters, through light welding at the seams between discrete passages, to the translation of large passages of new text with the aid of a term bank based on the organization's past usage. (*ibid*., 95)

In such a workstation (or, as he describes it, "a text-processing terminal with keyboard and screen"), there was still a role for a full MT system:

> Since this form of machine-assisted translation would operate in the context of a complete text-processing system, it could very conveniently be supplemented by "genuine" machine translation, perhaps to translate the missing areas in texts retrieved from the text memory. (*idem*.)

It was a concept not to come to fruition for another decade or more.

7. Martin Kay

One of the most decisive moments in the development of the future translator's workstation is now considered to be the (initially limited) circulation of a memorandum in 1980 by Martin Kay (1980). This combined a critique of the current approach to MT, namely the aim to produce systems which would essentially replace human translators or at best relegate them to post-editing and dictionary updating roles, and an argument for the development of translation tools which would actually be used by translators. Since this was before the development of microprocessors and personal computers, the context was a network of terminals linked to a mainframe computer. Kay's basic idea was that existing text-processing tools could be augmented incrementally with translation facilities. The basic need was a good multilingual text editor and a terminal with a split screen; to this would be added a facility to automatically look up any word or phrase in a dictionary, and the ability to refer to previous decisions by the translator to ensure consistency in translation; and finally to provide automatic translation of text segments, which the translator could opt to let the machine do without intervention and then post-edit the result, or which could be done interactively, i.e. the computer could ask the translator to resolve ambiguities. Here it is clear that Kay was building upon his previous work on the MIND system (Kay, 1973), where a 'disambiguator' would intervene during the analysis of a sentence to consult the human operator. Given a sentence such as (1), the computer might ask whether *tank* refers to a 'military vehicle' or to a 'vessel for fluids', whether *gas* refers to 'gasoline' or 'vapour', etc.

(1) They filled the tank with gas.

Kay stressed in his memorandum that he was advocating

> a view of the problem in which machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process... The keynote will be modesty. At each stage, we will do only what we know we can do reliably. (Kay 1980/1997: 13).

The result should be a system (a translator's *amanuensis*, as Kay called it) which "will always be under the tight control of a human translator. It is there to help increase his productivity and not to supplant him." (*ibid*., 20)

In retrospect, Kay's paper was more important for the force of its arguments against prevailing assumptions of MT researchers – in particular, the still residual belief that there were linguistic and computational answers (even if in the far future) to all translation problems and that full automation was a feasible objective – and for the force of its argument for the more modest aim of providing usable and acceptable translation aids for the working translator. (It is no criticism of Kay that he did not foresee the usability for many purposes of otherwise unacceptable computer-

generated translations. He was in good company for most of the 1980s.) What Kay did, therefore, was to provide the impetus for the development of a workstation that incrementally incorporated aids that were really wanted by professional translators. The fact that even his advocacy had relatively little effect until the late 1980s was a symptom of the continuing attraction of the full automation approach.

Kay's specific proposals included various means for translators to keep track of earlier decisions, whether relating to specific terms (for maintaining consistency), or to ideas of how to treat certain types of translation problems, or to previous occurrence of the same or similar passages, etc.:

> If the piece of text is anything but entirely straightforward, the translator might start by issuing a command causing the system to display anything in the store that might be relevant to it. This will bring to his attention decisions he made before the actual translation started, statistically significant words and phrases, and a record of anything that had attracted attention when it occurred before. Before going on, he can examine past and future fragments of text that contain similar material. (*ibid*., 19)

As well as this last conception of a kind of TM, Kay proposed a refinement of the usual replacement operation. He suggests that the translator should be able to insert special bracketing around problematic words or phrases:

> If it is, for the moment, unclear how a word or technical phrase should be treated, the tentative translation is enclosed in these special brackets… When the same word or phrase turns up again, the bracketed phrase is explicitly copied into the new position, thus maintaining an association among all the places where it is used. If the contents of such a pair of brackets are changed, the contents of all the others that are linked to it change automatically in the same way. (*idem*.)

The translator would 'bracket' only problematic instances, not all occurrences; thus, replacements would be selective. As a further refinement, Kay advocated the automatic generation of correct morphological forms for any "one of these bracketed words or phrases is written in a standard, regular form" (*idem*.)

8. Alan Melby

In 1981 Alan Melby put forward the use of a bilingual concordance as a valuable tool for translators. It enabled translators to identify text segments with potential translation equivalents in relevant contexts:

> [T]he source and translated texts are divided into units… 'translation segments'. For a given translation segment in the source text, the corresponding target language translation segment contains the translation of everything in the source segment, but there need be no decision as to the translation of a particular word. (Melby 1981a)

As an example, he showed an English text segmented into phrases (2a) and its corresponding French version, segmented likewise (2b)

(2) a. It was toward the close // of the fifth or sixth month of his seclusion, // and while the pestilence // raged most furiously abroad…"

b. Ce fut vers la fin // du cinquième ou sixième mois de sa retraite, // et pendant que le fléau // sévissent au dehors avec le plus de rage, …

The computer program would then create a concordance based on selected words or word pairs displaying words in context as in Figure1

more than one <at> a time	plus d'une à la fois
<At> first, as he spoke	D'abord, pendant qu'il parlait
<At> one time it was believed that	On a cru pendant longtemps que
meet <at> least once a year	siègent au moins une fois tous les 12 mois

Fig. 1. Concordance based on slected words

The concordance could be used not only as an aid to study and analyse translations, but also "for quickly determining whether or not a given term was translated consistently" (*ibid*., 545) in technical texts, to assist translators in lexical selection, and "in the development of a machine translation system for some narrow sublanguage." (*ibid.,* 546) Of course, the idea of generating concordances by computer goes back to the earliest non-numerical applications, but Melby seems to be the first to suggest their application as a translation tool. In his experiment, texts were input manually and correspondences between texts (later called 'alignments') were also made by human judgement. Only the concordancing program was automated, but Melby was clearly looking forward to the availability of electronically produced texts and of automatic 'alignment'.

At the same time, Melby was making specific proposals for a translator's workstation (Melby, 1981b; Melby, 1982; Melby, 1984) – made quite independently of Kay's proposals in 1980 (not widely distributed until some time later). During the 1970s, Melby had been involved in the MT research at Brigham Young University on a multilingual system with extensive human interaction during analysis stages (Melby, 1980). Experience on this project had sown the seeds for his ideas for a workstation of translation tools. A major problem encountered in the experimental system was that the translators involved were forced to answer uninteresting questions and to revise many sentences which they thought should be retranslated from scratch. It was a basic design assumption (shared by many MT systems of the time) that for economic reasons it was only worthwhile to use an MT system if the computer attempted to translate every sentence. This 'all-or-nothing' approach reduced the human translator or reviser to an unhappy 'garbage-collector.' Like Kay, Melby wanted the translator to be in control, to make his/her own decisions about when to translate fully and when to post-edit, and he wanted to assist translation from scratch by providing integrated computer aids.

The aim was the "smooth integration of human and machine translations" (Melby 1982: 217), bringing together various ideas for supporting translators in an environment offering three levels of assistance. At the first level, certain translation aids can be used without the source text having to be in machine-readable form. The translator could start by just typing in the translation as at a typewriter. This first level would be a text processor with integrated terminology aids and access to a bilingual terminology data bank, both in the form of a personal file of terms and in facilities for accessing remote termbanks (through telecommunications networks). In addition, there might be access at this level to a database of original and translated texts.

At the second level, the source text would be in machine-readable form. It would add a concordancing facility to find all occurrences of an unusual word or phrase in the text being translated, facilities to look up terms automatically in a local

term file, display possible translations, and means of automatically inserting selected terms into the text.

The third level would integrate the "translator work station"[1] with a "full-blown MT system" (as Arthern (1979) had also proposed.) Melby suggested that the ideal system would be one which evaluates the quality of its own output (from "probable human quality" to "deficient"), which the translator could choose to incorporate unchanged, to revise or to ignore. He pointed out that, although computationally and linguistically the third level was more complex than the second level, it would appear to the translator as very similar: whereas level two presented terms in sentences for evaluation and insertion, level three presented whole sentences for consideration.

Both Melby and Kay stressed the importance of allowing translators to use aids in ways they personally found most efficient. The difference between them was that whereas Melby proposed discrete levels of machine assistance, Kay proposed incremental augmentation of translator's computer-based facilities. Translators could increase their use of computer aids as and when they felt confident and satisfied with the results. And for both of them, full automation would play a part only if a MT system made for greater and cost-effective productivity.

9. From computer networks to stand-alone machines

These ideas of Kay and Melby were being made when text-processing systems still consisted essentially of a range of terminals connected to a mainframe computer and to separate printers for producing publishable final documents. It was natural to envisage networked systems rather than individual workstations.[2] For example, Melby assumed that the future scenario was a "distributed system in which each translator has a microcomputer tied into a loose network to share resources such as large dictionaries" (Melby 1982: 219). From today's perspective of cheap powerful personal computers it is instructive to cite Melby's description of the possibilities in the early 1980s:

> The individual translator work station would be a microcomputer with approximately 256K of main memory, dual diskette drives, CRT, keyboard, small printer, and communications port. Such systems are available at relatively low cost (under 5,000 U.S. dollars)... If several translators are in the same building, a local network can be set up to share terminology and document data bases and even inter-translator messages. The capabilities of the work station would include rapid, responsive word processing and access to internal dictionaries and to shared translator data bases... Access to source texts, document-specific dictionaries, and level three machine translations could be granted through a local network, a telecommunications network, or through the mails on diskette... (*idem*.)

At this date, even on-screen editing was a relatively new development. Nevertheless, translators of all kinds were quick to see the advantages of text processing—initially to reduce the amount of retyping involved in traditional methods. However, given the expense of text processing systems, the translation aids of the kind proposed were found mainly only in the larger translation services of major companies or in some translation agencies. The situation changed with the

[1] Melby's use of 'translator workstation' seems to be the first use of the term in its current meaning.
[2] At this date, the term 'workstation' in computer literature was used almost exclusively in the context of local area networks. The idea originated during the 1960s at Xerox Parc, where Martin Kay was employed.

appearance of the first personal computers in the mid 1980s (in particular the IBM XT and AT microcomputers), providing access to word processing and printing facilities within the range of individual professional translators.

Melby himself made the next significant advance with the provision of commercially available software for 'terminology management'– a component of level one of his design. A prototype was demonstrated in 1982 (Wright, 1988), and in 1987 the LinguaTech company was set up to market the Mercury (or Termex), later MTX, software package running on then current personal computers. It enabled translators to compile their own glossaries either as a separate task or while working on documents. The facility was designed to work interactively with a word processing program so that the translator could consult the glossary and transfer items to or from the text being translated. More importantly, links could be made to remote terminology databases and terms could be downloaded to the translator's own files, using Mercury's exchange format, MicroMATER (later developed into MARTIF.) In addition, the software could work in a network, so that a group of translators could share terminological data.

10. ALPS

Most significant for future developments were the facilities offered by ALPS (Automated Language Processing Systems) on its Multilingual Word Processor—programs running on an IBM AT personal computer—where the screen displayed source and target texts side by side; all formatting and typesetting codes were automatically copied from the source to the translated text without retyping; and there were facilities for copying any sections intact, e.g. tables and figures.

Firstly, the ALPS facilities included the AutoTerm software which enabled translators to obtain a glossary of terms for a specific text to be used during the translation process. Before starting, words of the source text would be matched against the stored dictionary, any not found could be added by the translator (in as much detail as desired) and the full list of terms formed a 'document dictionary' (or 'text-related glossary'). But whereas in the earlier Bundessprachenamt LEXIS system the list was printed out, in the ALPS system the terms were displayed in a 'reference window' at the bottom of the screen.

Secondly, ALPS provided a "repetitions processing" facility. Translators could specify that a 'repetitions file' be created as the translation process proceeded. As each segment of text was translated, it was copied with its translation into the file. Segments could be multiple-word phrases or longer syntactic units. Once in the file, the translator could compare a new text against files of segments from already translated texts, extract those that matched and then create a file of similar segments for use on the text currently being translated. The translator did not have to re-translate repeated parts of texts. This facility was called "repetitions extraction" (abbreviated as 'repstraction'), and was clearly an early implementation of basic features of a 'translation memory'.

AutoTerm and 'repstraction' formed the core facilities of ALPS systems. But Melby's idea of including optional automatic translation was also implemented. (Members of ALPS had been in the same research group as Melby at Brigham Young University, which developed a prototype interactive MT system.) The ALPS TransActive facility was an integrated MT engine which could be used by the translator to produce 'draft' translations of individual words or sentences; these could

be accepted or rejected, or form the basis for improvement interactively by human-computer dialogue. As in Kay's MIND system, users would be asked to resolve lexical and structural ambiguities. The designers were particularly concerned to minimise the repetition of interactions involving the same or very similar words and constructions. However achieved, the final translations could be edited on-line by the ALPS Translation Editor.

This combination of multilingual word processor, automatic term lookup, dictionary creation, document storage, access to previously translated segments, and interactive MT as required, formed a true forerunner of the translation workstation. ALPS itself was perhaps too soon on the market, since by the end of the decade the company was no longer selling its translation tool software, branching out instead as a translation service agency (ALPNET) – although continuing to use its own tools internally.

But ALPS was not alone. Elsewhere, there were other developments of workstations specifically designed for professional translators, which combined facilities for word processing, dictionaries, thesauri, split-screen display of source and target texts, and envisaging some access to automatic translation. They were intended not just for European languages, but there was also, for example, the workstation for English-Malay translation described by Tong (1987). Even so, they still lacked what translators would today regard as the 'core' component, the translation memory.

Translators knew already that a TM was what they wanted in an 'ideal' workstation – itself, by now, already a familiar concept (cf. Arthern, 1986). Not just good quality word processing, including non-English characters, good quality printing, customer-related glossaries, and easy retrieval of frequently used phrases, but also means of retrieving previous translations. As Vaumoron put it when listing requirements: the "source language text should be automatically coupled to the translated text, sentence by sentence." (Vaumoron 1988: 41). In other words, by the end of the 1980s, all components of today's translator workstation had been identified, but as a practicality it was still seen by translators as something for the future.

11. Bi-texts and statistical alignment

A seminal contribution towards the fully-fledged 'translation memory' concept was made by Brian Harris who put forward the "bi-text" as a translator's aid (Harris, 1988). He defined the bi-text as "a single text in two dimensions" (p.8), the source and target texts related by the activity of the translator through 'translation units' (i.e. it was essentially a refinement of Melby's earlier idea of a bilingual concordance). It is these linkages which the translator seeks to recall when working on a new text. The computer made it possible for translators to retain and retrieve these units:

> Now imagine a work station in which a translator has regularly stored all his translations with their originals in the form of bi-text, and together they make up a hypertext of several thousand pages... [If] the translator is stumped for the best translation of a certain ST word in the context... [the] 'search engine' is programmed in such a way that when it finds an occurrence of the word, it retrieves and displays the whole translation unit in which it occurs...
>
> [If] the translator needs help with... a translation unit [that] happens not to be a phrase that is conventional enough to have found a place in the dictionaries... we ask the search engine to look for... bi-text segments that will be similar enough to help him towards his objective...

> Thus bi-text used in this way could help the translator by providing... translations of words in context; a memory-perfect exploitation of the translator's own previous experience; near-translations of non-conventional phraseology and even longer units. (*ibid*., 9)

The final step of realisation was taken (ironically perhaps) by researchers in MT itself. At the same time as Harris was making his proposals, the DLT research group in Utrecht was formulating ideas on a 'bilingual knowledge base' (Sadler, 1989). In the DLT project, the proposed bilingual knowledge bank was to serve as a means of disambiguating source texts and generating appropriate target texts in the context of an interlingual MT system. The databank was to be constructed either from already existing bilingual corpora or in the course of interactive computer-based translation (as in the ALPS 'repstraction' facility). In either case, texts had to be analysed into relatively small 'translation units' (phrases, compound nouns) by means of the DLT parsers for both source and target languages. The interactive disambiguation of structures was as important to the construction of the bilingual database as it was for the translation system itself. However, although the DLT project itself made much use of statistical information as aids for automatic disambiguation, it was not proposed that the alignment of source and target units should be performed by statistical analysis.

For its antecedents, Sadler referred not just to Harris' 'bi-text' proposal but also to an earlier proposal by Makoto Nagao (1984) for an 'analogy-based' approach to MT – later to be known as 'example-based MT'. Nagao's basic argument was that translation is often a matter of finding or recalling how a particular source language expression or something similar to it had been translated before. A linguistic database of 'examples' would be derived from a structural analysis of a large corpus of source texts and their (human-produced) translations in the target language. The result would be sets of aligned bilingual phrases or segments.

It was, however, the development of statistical means for text alignment that made possible the realisation of these and earlier embryonic proposals for bilingual databases of translations. In the late 1980s, researchers at IBM reported their experiments on a statistics-based MT system (Brown et al., 1988, 1990). Systems for speech recognition and speech synthesis had been successfully developed with no reference to linguistic methods or theoretical constructs. Using the records of the Canadian Parliament (the Canadian Hansard) as a corpus of English and French texts, the IBM group investigated the application of their statistical methods to the task of translation. In brief, the method involved firstly the alignment of the corresponding English and French sentences and then the alignment within sentences of words (and phrases) which could be considered reciprocal translations.

In the subsequent years, there were and have continued to be many refinements and elaborations of programs for parallel text alignment (e.g. Brown et al. 1991; Church 1993; Dagan et al. 1993; Gale and Church, 1991; Isabelle et al. 1992; Kay and Röschenstein, 1993; Simard et al. 1992). Such statistical methods were exactly what were required to construct usable bilingual aligned databases for the 'translation memories' in translator's workstations. In practical implementation, their usefulness was dependent on the availability of large corpora of bilingual texts, specifically of source texts and of good-quality translations – and 'good quality' means usually translations produced by traditional non-MT means.

12. Translation workstations on the marketplace

During the 1980s, translation tools had become familiar to more and more translators. (For a good oversight of the situation see the articles collected by Vasconcellos, 1988). Further developments in microcomputing – the *de facto* 'standardization' of hardware platforms (IBM PC compatibles and Apple Macintoshs) and of user interfaces (MicroSoft Windows), and the appearance of high-capacity storage devices (internal and external) – accelerated the adoption of computer aids by the translation profession. They made possible the integration of translation tools in the form of translator workstations, which were actively developed from the end of the 1980s onwards, both public and commercial (Melby, 1992).

In the early 1990s, four commercial workstations for the professional translator appeared on the market at about the same time. The IBM Corporation launched its TranslationManager/2, the outcome of developments under Lippmann, after widespread internal testing. Other workstations appearing at the same time were the Transit system from STAR AG (a German company with many software products), the Eurolang Optimizer (in part a by-product of the GETA and Eurotra MT projects), and in particular the Translator's Workbench from Trados, a German company initially founded in 1987 as a translation service, which then developed translation tools from the INK company, and marketed the successful terminology management system MultiTerm in 1990. Trados was in fact the first to incorporate a 'translation memory' and alignment facilities into its workstation; and in recent years, it has added TAlign, a program enabling users to create their own translation memory files from existing translations. (It would appear that Trados was first to use the term 'translation memory' itself.)

Recently, commercial translator's workstations are integrating full-scale MT systems providing (as envisaged by Arthern, Melby, etc.) the option of sentence translations for consideration. Trados initially teamed up with Intergraph to incorporate the Transcend software[3], more recently it has provided an interface for the Logos system. In the case of IBM, the link has been with its own internally developed MT system, LMT; and Transit may offer Systran as its MT engine.

13. Concluding remarks

Research on translation tools continued throughout the early 1990s. The European Union funded the multinational TWB project (Kugler et al., 1995), and in Canada researchers from the previous Montreal MT project investigated a number of tools, e.g. the TransSearch bilingual concordancing tool providing more refined use of translation memories, and the TransCheck facility for identifying known common translation 'errors' such as deceptive cognates (English *actual*, French *actuel* 'current', English *physician*, French *physicien* 'physicist') and terminological inconsistencies (Isabelle et al., 1992). The translation service of the Commission of the EU is introducing an advanced workstation (EURAMIS) for its professional translators that integrates a multiplicity of multilingual tools, translation memories (both for individuals and groups), document databases (e.g. the CELEX database of EU legislation and directives), the vast lexical resources of Eurodicautom and of the Systran MT system, aids for technical writers (style checkers, drafting tools) and interfaces with publishing systems and to outside resources on the Internet (Theologitis, 1997).

[3] Subsequently, 'Transcend' was acquired by Transparent Language Inc.

In the future we might see an implementation of Kay's selective replacement idea. But, there is a particular need for matching phrases (as well as sentences) in translation memories and for composing retrieved fragments into coherent sentences. Current commercial translation memories are sentence-based and are restricted essentially to the presentation of potential examples of translations, which the translators must themselves adapt for incorporation. Searching for sentence fragments (phrases) is clearly beyond current statistical methods; it would require some linguistic analysis – as Melby, Harris and the DLT researchers assumed. Likewise, the automatic structuring of selected fragments into well-formed sentences demands a level of linguistic knowledge not yet available in commercial workstations.

The current enthusiasm among translators for the workstation is based on the successful integration of a variety of user-friendly and adaptable translation aids in ways that are particularly suitable for their own working practices. The workstation is the practical realisation of what Martin Kay and Alan Melby argued for in the early 1980s.

References

ALPAC: 1966, *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, DC.

Amano, S., Takeda, K. and Hasebe, K.: 1989, 'What should a translation work bench be like?', in *IFTT'89: harmonizing human beings and computers in translation. International forum for translation technology, April 26-28, 1989, Oiso*, JEIDA, Tokyo, pp. 15-18.

Arthern, P.J.: 1979, 'Machine translation and computerized terminology systems: a translator's viewpoint', in *Translating and the computer. Proceedings of a seminar, London, 14th November 1978*, ed. B.M.Snell, North-Holland, Amsterdam, pp. 77-108.

Arthern, P.: 1986, 'The translator's wordprocessor', *Language Monthly* 34, 24-25.

Boitet, C.: 1990, 'Towards personal MT: general design, dialogue structure, potential role of speech', in *Coling-90: papers presented to the 13th International Conference on Computational Linguistics… Helsinki*, ed. H.Karlgren, Helsinki University, Helsinki, vol. 3, pp. 30-35.

Brace, C., Vasconcellos, M. and Miller, L.C.: 1995, 'MT users and usage: Europe and the Americas', *MT News International* 12, 14-17.

Brown, P.F., J.Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin: 1988, 'A statistical approach to language translation', in *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics*, ed. D.Varga, John von Neumann Society for Computing Sciences, Budapest, pp. 71-76. Repr. in *Computational Linguistics* 16 (1990), 79-85.

Brown, P.F., J.Cocke, S.A. Della Pietra, V.J. Della Pietra, F. Jelinek, J.D. Lafferty, R.L. Mercer and P. Roossin: 1988, 'A statistical approach to language translation', *Computational Linguistics* 16, 79-85.

Brown, P.F., Lai, J.C. and Mercer, R.L.: 1991, 'Aligning sentences in parallel corpora', in *29th Annual Meeting of the Association for Computational Linguistics: proceedings of the conference, 1991 Jun 18, Berkeley, California,* [ACL], pp.169-176.

Church, K.W.: 1993, 'Char_align: a program for aligning parallel texts at the character level', in *31st Annual Meeting of the Association for Computational Linguistics: proceedings of the conference, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA.*, Association for Computational Linguistics, pp. 1-8

Church, K.W. and Hovy, E.: 1993, 'Good applications for crummy machine translation', *Machine Translation* 8 (4), 239-258

Dagan, I., Church, K.W. and Gale, W.A.: 1993, 'Robust bilingual word alignment for machine aided translation', in *Proceedings of the Workshop on Very large Corpora: Academic and Industrial Perspective, June 22, 1993, Ohio State University, Columbus, Ohio, USA,* Association for Computational Linguistics, pp. 1-8.

Flanagan, M.: 1996, 'Two years on – experiences, challenges, and trends', in *Expanding MT horizons: proceedings of the second conference of the Association for Machine Translation in the Americas, 2-5 October 1996, Montreal, Quebec, Canada*. [AMTA], pp. 192-197.

Gale, W. A. & Church, K.W.: 1991, 'A program for aligning sentences in bilingual corpora', in *29th Annual Meeting of the Association for Computational Linguistics: proceedings of the conference, 1991 Jun 18, Berkeley, California*, [ACL], pp. 177-184.

Gordon, I.: 1997, 'Translation memory systems', *ITI Conference 11... Proceedings of the Eleventh Conference of the Institute of Translation and Interpreting 8-10 May 1997… Harrogate*, ed. C. Greensmith and M. Vandamme, ITI, London, pp. 15-20.

Harris, B.: 1988, 'Bi-text, a new concept in translation theory', *Language Monthly* 54, 8-10.

Huang, X.: 1990, 'Machine translation in a monolingual environment', in *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June 1990, Linguistics Research Center, University of Texas at Austin*, LRC, Austin, TX., pp. 33-41.

Hutchins, W.J.: 1978, 'Machine translation and machine-aided translation', *Journal of Documentation* 34 (2), 119-159.

Isabelle, P. et al.: 1992, 'Translation analysis and translation automation', in *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, Empiricist vs Rationalist Methods in MT (TMI-92), June 25-27, 1992, Montréal, Canada*, CWARC, Laval (Québec), pp. 201-217.

Jacqmin, L.: 1992, 'La traduction automatique au service de l'utilisateur monolingue', *Meta* 37 (4), 610-623.

Jones, D. and Tsujii, J.I.: 1990, 'Interactive high quality translation for monolinguals', in *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June 1990, Linguistics Research Center, University of Texas at Austin,* LRC, Austin, TX., pp. 43-46

Kay, M.: 1973, 'The MIND system', in *Natural language processing: Courant Computer Science Symposium no.8, 20-21 December 1971*, ed. R. Rustin, Algorithmics Press, New York, pp. 155-188.

Kay, M.: 1980, *The proper place of men and machines in language translation*. Research report CSL-80-11, Xerox Palo Alto Research Center, Palo Alto, CA. Repr. in *Machine Translation* 12 (1/2), 1997, 3-23

Kay, M. and Roschenstein, M.: 1993, 'Text-translation alignment', *Computational Linguistics* 19 (1), 121-142.

Krollmann, F.: 1971, 'Linguistic data banks and the technical translator', *Meta* 16, 117-124.

Krollmann, F.: 1974, 'Data processing at the translator's service', *Babel* 20, 121-129.

Krollmann, F., Schuck, H.J. and Winkler, U.: 1965, 'Herstellung textbezogener Fachwortlisten mit einem Digitalrechner: ein Verfahren der automatischen Übersetzungshilfe', *Beiträge zur Sprachkunde und Informationsverarbeitung* 5, 7-30.

Kugler, M., Ahmad, K. and Thurmair, G. (eds.): 1995, *Translator's workbench: tools and terminology for translation and text processing*, Springer, Berlin. (Research Reports ESPRIT Project 2315).

Lippmann, E.O.: 1971, 'An approach to computer-aided translation', *IEEE Transactions on Engineering Writing and Speech* 14 (1), 10-33.

Lippmann, E.O.: 1975, 'On-line generation of terminological digests in language translation. An aid in terminology processing', *IEEE Transactions on Professional Communication* 18 (4), 309-319.

Lippmann, E.O. and Plath, W.J.: 1970, 'Time-sharing and computer-aided translation', *The Finite String* 7(8), 1-4

Melby, A.K.: 1980, 'ITS: interactive translation system', in: *Coling 80: Proceedings of the Eighth International Conference on Computational Linguistics,* Tokyo, Japan, pp. 424-429

Melby, A.K.: 1981a, 'A bilingual concordance system and its use in linguistic studies', in: *The Eighth LACUS Forum 1981,* ed. W. Gutwinski and G.Jolly, Hornbeam Press, Columbia, SC. (1982), pp. 541-549.

Melby, A.K.: 1981b, 'Translators and machines – can they cooperate?' *Meta* 26(1), 23-34.

Melby, A.K.: 1982, 'Multi-level translation aids in a distributed system', in *Coling 82. Proceedings of the Ninth International Conference on Computational Linguistics, Prague, July 5-10, 1982*, ed. J.Horecký, North-Holland, Amsterdam, pp. 215-220.

Melby, A.K.: 1984, 'Machine translation with post editing versus a three-level integrated translator aid system' [presented at International Conference on Machine Translation, Cranfield University,

13-15 February 1984], in *Progress in (machine translation: natural language and personal computers,* ed. I.D.K.Kelly, Sigma Press*,* Wilmslow, UK. (1989), pp. 119-125.

Melby, A.K.: 1992, 'The translator workstation', in *Computers in translation: a practical appraisal*, ed. J.Newton, Routledge, London, pp. 147-165.

Nagao, M.: 1984, 'A framework of a mechanical translation between Japanese and English by analogy principle', in *Artificial and human intelligence*, ed. A.Elithorn and R.Banerji, North-Holland, Amsterdam, pp. 173-180.

Nirenburg, S.: 1992, 'Tools for machine-aided translation: the CMU TWS', *Meta* 37 (4), 709-720.

Oettinger, A.G.: 1960, *Automatic language translation: lexical and technical aspects, with particular reference to Russian,* Harvard University Press, Cambridge, Mass.

Sadler, V.: 1989, *Working with analogical semantics: disambiguation techniques in DLT*, Foris, Dordrecht.

Simard, M. et al.: 1992, 'Using cognates to align sentences in bilingual corpora', in *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, Empiricist vs. Rationalist Methods in MT (TMI-92), June 25-27, 1992, Montréal, Canada*, CWARC, Laval (Québec), pp. 67-81.

Somers, H., Tsujii, J.I. and Jones, D.: 1990, 'Machine translation without a source text', in *Coling-90: papers presented to the 13th International Conference on Computational Linguistics… Helsinki*, ed. H.Karlgren, Helsinki University, Helsinki, vol. 3, pp. 271-276.

Theologitis, D.: 1997, 'EURAMIS, the platform of the EC translator', in *1997 EAMT Workshop, Copenhagen, 21-22 May 1997, Proceedings: Language technology in your organization?* Center for Sprogteknologi, Copenhagen, pp. 17-32.

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

Vasconcellos, M. (ed.): 1988, *Technology as translation strategy*, State University of New York Press, Binghamton, NY. (American Translators Association Scholarly Monograph Series vol.2)

Vaumoron, J.A.: 1988, 'The ideal workstation? Perspective of the in-house translator', in Vasconcellos (1988), pp. 36-41.

Wright, L.D.: 1988, 'Mercury: a tool for the management of specialized terminology', in Vasconcellos (1988), pp. 92-96