
Report

Ten years of Translating and the Computer

The tenth annual Translating and the Computer conference, undoubtedly one of the major events in the language world, was held in London on November 10 and 11, 1988. Peter Cholas reports.

It was generally agreed by participants that this was one of the better conferences in this remarkable series. There did seem to be a prevailing impression that there was a lull in new developments in machine translation, or at least the public announcement of them, but that there was considerable dynamism in the actual application of MT.

The conference was opened by Professor Juan Sager, with the survey of the state of the art which is now customary on these occasions. He recalled that translators, at first suspicious of the new technology ("translators are traditionalists by profession"), now showed interest in developments, while producers of systems have learned to sell their products without extravagant claims. "We now have information; before we only had publicity", he said.

A good rapport had been established in the design of systems. The designers had learned to listen and talk to translators, and had found that the translation process is more complex than had earlier been conceived. There had been a realisation of the need for models of translation, which must have a theoretical basis.

The present situation was that translators knew of the need to work with machine input and output, while research in MT was now moving in the right direction, into applications. Translation, whether human translation or machine translation, could not be separated from document produc-

tion. There was a need for systems which took into account broad textual features, and could understand things not necessarily explicitly stated in the text. System designers must be provided with translation specifications. One area of MT application would be for non-linguist readers of technical and scientific literature. Finally Professor Sager lamented the shortage of people working in MT design.

Fred Zirkle, president and chief executive officer of ALPS and ALPNET, gave a frank account of how his company was adapting to the changing needs of the translation market. There was no question of forsaking system development for the translation services scene. ALPS had spent \$24 million already on machine translation system design, and they had an ongoing programme of software development. These days the gulf between the software developer and translators was fast disappearing.

He indicated that all firms marketing machine translation systems were having difficulty making a profit; "all", he said, "are showing a retained earnings deficit." He guessed that the total revenue would be about one half of the money spent on developing systems, and companies in this field had carried out significant staff reductions this last year.

But machine aided translation had been shown, he claimed, to increase throughput of work fivefold, and with its other advantage of consistent quality was finding growing acceptance. "We feel computer aided translation is inevitable", he said, and he felt that the tide had now turned towards it.

Mr Zirkle then turned his attention to the translation business, and the

need for reform. The pattern in the past had been a range of large multinational customers, and small regional suppliers; the average translation company had six employees. Yet the top 50 users of translation services would need the resources of 500 typical suppliers.

Although, as he had emphasised, ALPS was not getting out of software development, a technology-driven company of its type had the problem of the high cost of sales, of training, and of customer support, contrasted with a low basis of installed systems, and low profits. "Becoming our own user as well as vendor" therefore made commercial sense.

Customers who were encouraged to contract with ALPS for services may later purchase the software, and in any case the development team benefited from useful feedback. So ALPS, which through its purchases of translation companies now had nine translation firms with 22 offices throughout the world, was "de-emphasising" sales of its systems where there were less than three work stations.

Mr Zirkle forecast that "a new breed of translation supplier" was bound to arise. Instead of the possibly competent but small and scattered translation company of the past, there would arise companies operating on a worldwide basis, with vertical integration, large in-house staffs, and "global project management"; translation companies, in fact, which "look like the customer".

Such developments could only enhance the status of the translator, he said. Tomorrow's translator would be "computer-comfortable", and an employee rather than a freelance.

"This is an exciting time to be involved in translation. Demand for quality has grown exponentially."

A paper on an attempt to apply machine translation in a smaller environment was given by *Isabella Moore*, who described a project at Aston University. She recounted many of the difficulties and pitfalls, and observed how heavily dependent on the supplier and

the quality of his support the small user became.

David Jackson, managing director of Vuman Computer Systems Limited, which developed a word processing package for translators, described the progress made over the years in providing accented characters in word processing programs. He speculated that the new IBM P52, with its "Code Page 850", might impose itself as an industry standard.

The successful application of a machine translation system in an industrial firm was described by *Peter Pym*, technical publications manager for Perkins Engines Limited, of Peterborough, UK. The key, he made clear, was that the input text was always written in a form of Controlled English.

This system, called PACE (Perkins Approved Clear English), had been introduced into the firm in 1980. Its lineage could be traced back to a controlled English system marketed as ILSAM (International Language for Service and Maintenance), which in turn derived from Caterpillar Fundamental English, originally developed in the United States by the Caterpillar Corporation.

The golden rule for establishing PACE words was "one word, one meaning". For example the word *right* is often used in speech to mean "correct". In PACE *right* is defined as "when facing north, right is east", while *correct* is used in the sense of "conforming to a standard, opposite of wrong".

One of the ten rules for simplified writing in PACE was the avoidance of strings of nouns, such as "the cast iron timing case cover". The -ing ending, or gerund, was avoided since it often leads to ambiguity. In PACE one tried to get the length of the sentence down to about 20 words, and preferably about 15.

Mr Pym gave a number of examples of the way PACE affected technical writing, including the following.

Pre-PACE

The heavy duty oil bath air cleaners are usually fitted with a centrifugal pre-cleaner

mounted on top of the main cleaner. This should be removed and the air inlet vanes in the bottom plate of the assembly, the ejection slots on the side of the cone and the vanes in the outlet tube, cleaned of dust and dirt.

PACE version

Heavy-duty air cleaners of the oil bath type are usually fitted with a centrifugal pre-cleaner, which is mounted on top of the main cleaner. Remove the pre-cleaner. Clean completely the air inlet vanes in the bottom plate of the assembly, the ejection slots on the side of the cone and the vanes in the outlet tube.

Up to 1987 translations for Perkins, he said, had been produced mainly by the company's local distributors. The translations and the format of the publications were inconsistent, and they stretched local resources. His department, which consisted of engineers, not linguists, had been given the brief to produce good translations quickly for new ranges of products, but without additional staff.

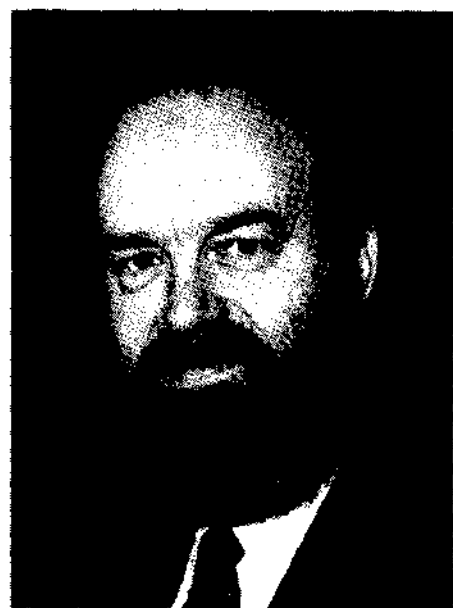
The company had looked at the Weidner MicroCAT system in 1984, and it had been acquired for trial in 1985. Four PCs had been acquired for its use in 1986 and 1987. University lecturer *Tony Hartley* had been brought in to advise on translation quality.

PACE had been introduced and expanded in a controlled manner. Approximately 2,500 words, of which 250 verbs, are used in the writing. The objective was to produce text which was clear, succinct and readable. The translations produced by the system could be post-edited at a speed of between 7,000 and 9,000 words a day, and translation costs had been cut by 50%.

Another example of the application of machine translation, and of the integration of translation into the general document preparation process, was given by *Mike Scott*, of Rank Xerox, who revealed that the typical desktop copier had a "life" of only 18 months before the competition would catch up. Therefore getting a product to the market simultaneously in all countries was a vital necessity. If translation was on the "critical path" of the development schedule, then considerable pressure was put to produce the translation quickly. Rank Xerox used both the Systran and the ALPS systems.



The final session of the conference was planned as a tribute to the late *Margaret Masterman*, one of the leading figures in early MT research. *Tony Hartley*, who chaired the session, recalled that she had always asked for a translation model, how the translator works.



Professor Frank Knowles recalled that for *Margaret Masterman*, "human translation is the acme of skilled linguistic activity", possibly rating higher than creative writing, since the translator has to reveal the brilliance of the author while concealing the brilliance of his own!. She wanted machine translation to emulate the behaviour of the human translator; nothing less would

do.

Professor Knowles then gave a tightly argued paper on how machine aids could help translators. Machine aids could take many forms- facsimile and OCR, for example, could be seen as examples of bit-mapping. There was the problem of the archiving of translated documents, whether simply with ASCII codes, which might mean abandoning macrostructural information, or by some means which preserved the format. Even software aspects which seemed at first of little relevance to the translator, such as spread sheets, could be used for varying mathematical data, such as for conversion of the different tyre pressure units used in different countries when translating automotive documents. Database facilities had been much discussed, but ought to include domain tags and cross-referencing provision. Better dictionaries might become available on CD-ROM, though there was of course the danger of their perpetuating the deficiencies of existing dictionaries — “we must hope that flexible database management software will soon be applied to materials of this sort, thereby offering us all a quantum leap in intellectual, functional and operational terms”.

There was a need for devices which helped translators to “boilerplate” translations, by for example the incorporation of constantly repeated legal phrases.

Ranging more and more widely, Professor Knowles meditated on translators’ need for de-hyphenation software, multilingual spell-check software, and software to highlight all potential technical terms, based on the stratagem that the longest match offers the highest probability of correct identification.

He also considered the possibility of a utility showing a one-to-many relationship between source and target lexical items: “a very simple concordance, which could fairly easily be enlarged as a by-product of the ongoing translation process, needs to be generated: for each source item — in English, say — a 100-example batch,

say, of contextually arrayed German equivalences could be retrieved for inspection and could thus help to prime a judicious choice”.

Another device useful to the translator was word-frequency information, but what was of great importance was access to specialised information.

“The on-line consultation of domain thesauri is another highly desirable facility: there is a need for translators to have available various technical terminologies presented in an onomasiological fashion, reflecting real-life linkages rather than the vagaries inherent in alphabetically organised compendia. Alphabetic entry into such subject microcosms is, of course, a welcome convenience — but the fundamental value of such thesauri is to verify knowledge structures and context. The ideal case, of course, is a fully co-ordinated multilingual reference system, encyclopaedically arranged and incorporating copious illustration. Unfortunately, such works are not yet fully available — even in traditional printing — either as professional or pedagogical compendia. What little does exist offers precious little in terms of amenability for translating purposes. Only computerisation can release the full power of such a concept — translators must encourage such

developments and, preferably, participate in them”.

Turning his attention to the hardware Professor Knowles sketched a concept of the “translator’s work station”. This would include a 32-bit processor with 1 Mb of RAM backed up by 100 Mb of spinning store which was in turn backed up by tape-streamed archives. The screen should be high-resolution colour and of A3 size so as to allow synoptic viewing and handling of enlarged text and graphics as well as of multiple windows.

More thought needed to be given to the division of labour between person and machine, with pre-editing, intra-editing and post-editing playing their part (“human beings saving the machine from itself”).

Professor Knowles looked at the EUROTRA project of the European Commission, and regretted that more information, and more openness about the progress or lack of progress, was not made available. There should be no need for casuistry, but this is what

emerged when there was a mismatch of expectations in the triangle of contractor, supplier and customer. The last thing we need now in Europe is another ALPAC-like report, he said.* In MT and MAT what we needed was less politicking, and more *glasnost*. Indeed generally more frankness was needed about the problems which still remain to be resolved in MT.

In a major passage which is likely to reverberate in future discussions of the progress of EUROTRA, Professor Knowles obliquely criticised past approaches, and tried to view the way forward.

Only bureaucratic error or political disingenuousness could lay down in advance in 'blue skies' research that a "research phase" WILL be complete in two years. Only an honest assessment of work achieved — rather than "milestones" specified *a priori* — can determine progress and the managerial options surrounding it....

...If a large R&D effort is to be distributed across several centres then a proper assessment of the contingent financial, logistic and 'personal chemistry' overheads should be conducted. If it is proposed to 'add' languages to developing systems, what evaluation of start-up effort is even attempted? Is it easier to lexicographically codify the first 5,000 words or the last 5% in a restricted vocabulary? Cannot the aggregation — even at a

*The ALPAC report in the United States in the early days of machine translation decided on the evidence of some unsuccessful experiments that machine translation could not succeed, and produced a major setback to development in the field.

seemingly glacial pace — of lexicographical material provoke new insights into lexical "structurality"?

The point of these largely rhetorical questions is really to suggest that developing a multilingual, multi-functional system such as EUROTRA is really akin to NASA trying to launch a space-shot to Mars. The true nature of the complexity is not known beforehand and those funding it having to accept that quite openly at the beginning...

Future developments in MT and MAT, opined Professor Knowles, would come not from loners but from members of closely-knit teams, and he hoped that co-operation would spread.

After all, lots of customisation will be involved, lots of market and product research will be needed, lots of financial shrewdness will be required to cope with shortening timescales for equipment depreciation, with continually enhanceable software and with 'add-on' language pairs.

It is to be hoped that a coalescence of interests, firstly among providers, and subsequently extending into the customer-base for MT and MAT, will make it possible to integrate effort, avoid duplication, set pragmatic but professional standards and serve to create a climate in which MT and MAT research and development will be seen by all parties not as expenditure but as an indispensable investment.

A different aspect of language activity was examined by *Dr Georg-Michael Luyken*, deputy director of the European Institute for the Media, who considered the language implications of the revolution in television programming with the advent of satellite television, which he described as the first

profound revolution in the medium since it began. It would lead to growing internationalisation (more co-production deals), and growing commercialisation.

Dr Luyken described the different systems of dubbing (preferred in Germany, France, Italy and Spain), sub-titling (preferred in the Netherlands and Scandinavia) and voice-over, and mentioned that it was generally reckoned that a programme needed to be receivable in 20 million households to be worth the cost of dubbing. The cost was between £10,000 and £20,000 for each hour dubbed, including the hire of the dubbing theatre, the raw translation of the script (usually done by "low-cost suppliers"), the adaptation of the script to lip movements on the screen under the supervision of the dubbing director, the actors, and the track mix. Sub-titling cost only 1/15th of the costs of dubbing, but a reduction in content was inevitable. He forecast there was going to be a big demand for language conversion in programmes, with dubbing remaining the main medium, but with more and more sub-titling, especially for programmes with fringe interest.

The final speaker at the conference, *Professor Yorick Wilks*, spoke directly about Margaret Masterman's work, and speculated on what would be her attitude to many current developments.