ASLIB/TG CONFERENCE

"TRANSLATING AND THE COMPUTER 6: TRANSLATION AND COMMUNICATION"

First day - November 20th

Birgit Rommel, Dolmetscherschule, Zürich, Switzerland, spoke on new approaches to training. Zürich University trains translation students to use word processors and word processing packages and to create computer awareness in its students. To this end, members of staff have joined training programmes run by the University's Computer Centre. In class, the lecturers work at terminals connected to the main computer and the students observe the work in progress via monitors. An insight is provided into the fact that not all word processors/packages are suitable for multilingual work, and this from various points of view. The philosophy is that students may be required to advise their future employers on systems suitable for translation work, or indeed have to order their own equipment if they become freelances. On balance Zürich University favours the dedicated word processors, these being faster and more comprehensive than word processing packages running on micros.

Lisa Price, Centre for Computational Linguistics, UMIST, Manchester, UK, spoke on the British Term Bank Prototype. She discussed the database structure, the requirements of the intended user-groups, and the human aspects of accessing and using a term bank online. Recognizing the varying levels of user experience and needs, a flexible menu-based interface has been designed, and the prototype will be demonstrated early in 1985. An outline was provided of future work on the following: theoretical terminological research (eg on word compounds), the construction of a translator's work-station (allowing split screen, simultaneous accessing and word processing), and a user study to determine the most effective and feasible modes of operation in the UK. The data base management system is the nucleus of the prototype; there are connections between the files which are subdivided according to language. The existing model permits access via English only, however there are other indices in other languages. The system is user-friendly, avoids accidental log-off and has help features. Entry is by user-name and password, the user defining the output format at this point: the output can be tailored to the user's requirements. Searches on character strings can be made of the whole database or a subset of same. Various truncation possibilities also exist, although these may lead to unwanted terms being provided by the term bank.

Bill Medcraft, Specialist Services, British Telecom, London, UK spoke on international communications. He noted that text and data services in all countries are shaped by terminology, standards, protocols and the changing needs and requirements of users. In terms of technology, powerful desktop micros are now available, and telephones and exchanges are now essentially specialized computers.

14

Transmission systems now include such devices as satellites and optical fibres (which may carry up to 1700 simultaneous calls per fibre), together with digital operation to limit signal degradation. Digitization is used for the purposes of easy transmission and switching. The packet switching system permits text messages to be broken into blocks containing a header and a tail: all such blocks travel over the system in an independent manner. International bodies such as the CCITT, CEPT and ISO lay down standards in these areas. For example, the CCITT X25 standard covers the connection of micros to the public data networks. International communications require the observance of procedures between national networks and there is a CCITT standard for linking electronic mailboxes. All such standards must fit together to form an interlocking jigsaw. However the international telex system is crude and teletex offers high-quality text transmission at 50 times telex speed. In terms of user needs, the perceived value of information is increasing: on an international basis text transmission services overcome time, national and language barriers. Public service networks and private circuits are available for communication both between and within companies. The best UK answer in this area is the Packet SwitchStream service which supports terminals ranging from simple teletypes to mainframe computers, and acts as a gateway to the international packet switching system (which basically covers the northern hemisphere western countries plus Japan, Singapore, Australia, etc.). Using this system the transmission costs are a function of the quantity of copy sent and not the transmission time.

Peter Wheeler, Logos Computer Systems Deutschland GmbH, Frankfurt, FRG, provided an update on Machine Translation(MT). He noted that German has risen to prominence as both SL and TL, eg the Logos system is being evaluated in the EC in Luxembourg on a six-month trial basis and Systran is involved in a German-English joint venture with the USAF, the World Translation Center and Festo. The University of Saarbrücken now has Betsy, a lexicon containing complex technical noun groups only. There is an accelerating understanding of the need for both high-speed and not very elegant translation side-by-side with translation of the highest quality. MT Systems have more frequent recourse to rerunning failed translations under less strict grammatical parameters, on the basis of any translation being better than none. There is still activity in French, eg at Grenoble, where work is also being carried out on English to Malay and to Thai. The French Ariane 78 system has been selected as the basis for the French Government's National Machine Translation Project: English to French and vice-versa. Eurotra is proceeding, the complete date for a 20,000-word prototype being mid-1988 to be executed by 10 national teams participating in one overall project: the experimental software assembly is ready and the grammar writing will start in January 1985. There has been much activity in Japan, eg the Japanese Fifth Generation Project which is aiming for informative-level translation of scientific and technical abstracts, both English-Japanese and vice-versa. Several commercial projects have come onto the Japanese market during 1984 and Weidner has also brought out a Japanese package. Commercially, MT is now in the process of being regarded as any other investment project in business would be. Many who were potential users in 1983 are still looking - their reasons being largely the non-availability of the appropriate language combinations, non-matching

hardware configurations, and insufficiently high quality. Spanish is a language which is also active: General Motors, Canada, is machine translating truck and car manuals and PAHO is using Engspan on a pilot project. In other areas Alps has added both Russian and Arabic facilities together with other European languages at one level or another in their system, and have set up a regional office in Switzerland. The Weidner system at ITT Europe has been so successful on worldwide internal company translation that ITT are taking on contract work from outside and increasing their staff; such increases also being reported from other MT units. Systran has added English to Arabic.

<div align="right">Michael Fulton</div>

Parallel sessions on the state of the Art
----

1 - Handling terminology

Barbara Wilson, (Translation Branch, Foreign and Commonwealth Office) chaired the lively informal discussion, introducing Kees van der Horst of the Terminology Department, Commission of the European Communities (EC), Luxembourg. He is responsible for Eurodicautom, a term bank with the stupendous total of 300,000 English terms as well as their translations into other Community languages and references to clarify their meaning.

This vast stock of text is accommodated on 200 megabytes of hard magnetic disk (averaging 700 characters per English term) for the database, and 400 kilobytes for the program.

Eurodicautom, having more means of access than the purely alphabetical system of card files, is a more fruitful way of looking for the translation of a term. The system is used by EC translators and is available on Euronet-Diane through ECHO, the European Community Host Organization, which offers access to selected EC databases free of charge and caters especially for the small user and those new to online searching.

Sarah Dunlop, of ECHO, Luxembourg, Manager for Education and Training (of users) explained the accessing method. The best way from the UK is to use British Telecom's PSS (packet switching service). This cheap-rate data transmission network allows a user to connect his terminal or microcomputer with the ECHO computer in Luxembourg.

To make a connection, the user dials the nearest PSS pick-up point (node). When the call is accepted he identifies himself to the network by keying in his password (NUI, obtained from PSS) and then keys in the ECHO call number (NUA) to tell the network which remote computer he wishes to be connnected to. Upon arrival, the user is requested by ECHO's computer to key in his user password (supplied by ECHO free of charge) to gain access to Eurodicautom.

The user's costs for accessing ECHO are limited to (1) a local telephone call, (2) PSS charges of about £6.25 per quarter plus £1 per connect hour, and a once-only password fee of £25.

The sequence can be shown diagrammatically as follows; (LUXPAC is the Luxembourg data transmission network).
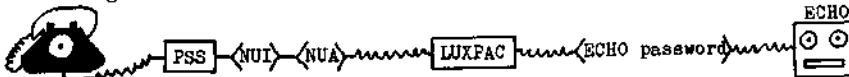
A communications package (the programs needed for getting one's micro to communicate with the Eurodicautom computer) can be obtained without difficulty. In fact ECHO can supply such a program for the BBC micro on request. Sarah Dunlop demonstrated on her micro how to do this, obtaining text from Eurodicautom on the screen.

Kees van der Horst explained that much of the Eurodicautom text consists of glossaries from the British Standards Institution, the International Electrotechnical Committee or the French firm Aerospatiale. He was always willing to accept new glossaries, especially the obscure ones, and there was a budget to pay for them. Each term had its 'subject code' to indicate its general category and the computer could be made to pick out all terms with this code and print them as a specialized glossary.

ECHO has some 1400 users, more than half of whom are inquirers into Eurodicautom, and some of these are freelance translators.

Several speakers told us of their experiences, including Helga Liebgott of Ruhrgas AG, Essen, and Patrick Sutton of the Department of Health and Social Security, whose glossary of social security terms will include French and German equivalents and should have 500 terms in each language when published. Patrick Chaffey said that the ADNOM glossary with its standard English equivalents of translations of Norwegian official bodies would soon appear.

The session was intended to be open for debate on any aspect of terminology but despite the efforts of the chair to broaden the scope of the lively discussion it kept returning to Eurodicautom, reflecting the interest of the audience in a most accessible databank.

John Scott

2 - Translation aids for the small translation unit

Several main areas were discussed. The first was on input and the cost-effectiveness of dictation coupled with audio-typing versus direct keyboarding onto word processing systems: the former seems dependent on cheap, efficient typists who are not to be found in all countries. The consensus was that the best method was really a function of individual experience and preference. The second topic was hardware: the discussion centred around the difficulties posed by restricted character sets on the one hand, and the need for letter quality dot-matrix printers, disk compatibility, and A4-capacity screens on the other. It was felt that translators needed guidance in this area and that manufacturers should be more aware of the problems: several contributors reported that both their translation quality and quantity improved enormously through word processing. The final section was on output: there were adherents both of editing on-screen and hard copy editing. It was also found that many translators are already using telephone transmission of copy; the shortcomings of facsimile quality were also mentioned.

Bill Duffin

3 - Machine translation

Linn Downs, of Burroughs, provided the session with a set of criteria for selecting an MT vendor: viz. target the problem, appoint a team (to contain both financial and a systems analyst), investigate and compare MT software vendors, prepare an installation schedule using critical path analysis and make the decision. The subsequent general discussion dealt with many topics. With regard to MT output quality there were 2 points of view, viz. that all translation should be top-quality, and the opinion that the market sometimes requires a faster cheaper not-so-polished product purely for information purposes. There was discussion on the post-editing of MT translations required to obtain a top quality product: post-editors need to be translators with subject familiarity, word processing skills and good motivation, this latter being of great importance. Experience shows that they effect less revision with

time and that the vocabulary within the original SL document may
need to be more consistent for successful MT. Word processing and
MT can now be integrated with publishing systems thus leading to a
reduction in proofreading problems. It was felt that the various
claims for percentage acceptability of the MT product are open to
differing interpretation. There are different types of MT: some
MT systems run overnight, with editing off-line but this is not
possible with the interactive type of MT system. Interactive MT
permits making changes during the translation process and tailoring
the grammatical capabilities. It is clear that the slowest step
in MT is often the manual input (keying-in) and that economical
desk-top OCR (optical character recognition) units are required
which will accept the diacritical marks found in a great many Roman
character set languages. We were told that Eurotra should cover
all EC language pairs, albeit with a limited vocabulary initially -
it would thus be more powerful in the sense that it will handle
pairs of minority languages. It was felt that it may be inappro-
priate to produce Arabic, Farsi, Japanese, etc. MT packages as by-
products of West European language programs due to the fundamental
differences between such language groups.

<div align="right">Michael Fulton</div>

Second day - November 21st

Wednesday proved a hard day for many of us as we grappled with
jargon from a fascinating variety of branches of electronics. We
began easily enough with Linda Talhami describing the effects of
office automation (OA) on the language services provided by Tele-
globe Canada. This government-financed corporation, located in
Montreal, provides translation and terminological services which
are available to companies and individuals as well as government
departments in a country where bilingualism is official policy.
Teleglobe, with around 900 translators, is large enough for OA to
have had significant effects on productivity over the 5 or 6 years
since its first appearance.

In the initial stages, the economic impact was small because OA
was based on linkage to a main-frame computer through operators,
while translators still worked with pen and typewriter. During
the last 3 years, however, terminologists and translators have
been using terminals directly and of the 45% increase in product-
ivity recorded over that period, 20% is reckoned to be due to
word-processing facilities. A detailed description of the stages
needed to produce a finished translation showed that the 7 hard
copy versions needed before OA had been reduced to 3 afterwards.

The advantages of OA were listed as time-saving, improved product-
ivity, efficiency and quality, and increased morale. Experience
showed that there were obstacles to its introduction, notably a
resistance to change by staff (best overcome by a relatively slow
pace of development and working as a team). There was, however, a
wider aspect to OA: the increased breadth and efficiency of the
language services that could be offered, including creation and
up-dating of glossaries, index generation, handling of statistics,
and terminological work including spelling checks.

The survey concluded by summarizing the main features in intro-

ducing OA as (a) identification of applications, (b) social issues such as increases or decreases in motivation, (c) its role in raising productivity, not least because both translators and clients were more satisfied with the results.

Robert Rooke of Textnet then took us forward to a further stage of communication, that of Electronic Mail. In its original form, the sending of text over the telephone network cost thousands to set up and needed a large telephone exchange to install it. Today the equipment is portable (even to a public call box) and the initial cost is only in the hundreds. In one form of what he called 'point-to-point' communication, where the two endpoints are micro-computers, all that is needed at each end is a modem: a device that encodes the micro output into a form suitable for sending down a telephone line or decodes it at the other end for the receiving micro. This system can operate at speeds of some 1000 English words every 3.3 minutes at a moderate cost.

A new development known as packet switching makes possible trans-mission over larger distances at higher speeds - and more cheaply. Moreover, unlike the situation with direct dialling, data cannot be lost. Speeds of transmission are around 6400 characters per second and a typical cost for London-Paris would be 40p per 1000 words (compared with £2.20 per 1000 using direct dialling). With this method, the user accesses a central computer which can act as an electronic mailbox in holding incoming and outgoing text until called, and can even inform senders if their letters have been read by the addressees! This network can be accessed from hotel rooms or callboxes so that a translator's work need no longer be dependent on time or distance.

After a coffee break which allowed us time to reflect on the various paradises now being revealed, we returned to a different format for the last morning session. Four major machine translation (MT) interests - ALPS, Logos, Systran and Weidner - were represented on a panel and, after a brief introductory statement from each, were exposed to questions and comment from the floor. Since information on the capabilities of these systems is available from the firms themselves, there is little point in covering that ground here. Merle Tenney of ALPS and Ralph Hawes of Logos emphasized opposite aspects of their products: the first being highly interactive and the second fully automatic. Dr F. Habermann, a user of Systran rather than a company representative, gave figures for the increase in comprehensibility of output over the last 4 years (from 94% to 98% no less!) and Logos mentioned the same point. Stephen Kerce for Weidner emphasized their system as 'available to small organizations'.

Questions from the floor were restated by the Chair, Catriona Picken, ostensibly so that everybody could hear, but her tactful and expert rephrasing added much point to what was being asked. In fact, little chance was taken to challenge the claims of the MT systems, perhaps because we are now more realistic about what they are capable of doing and what they are not. The reported improve-ments in comprehensibility were queried as to their objectivity and independence but this only elicited detail about the tests carried out. A query about semantic competence brought out the

fact that Logos uses 32,000 rules in semantic tables (a figure that may increase to 100k) and this copes with about 90% of text!

The possibility of marketing MT for small users having need of only one language pair brought some hope from ALPS and Logos that the increasing power of personal computers might mean that something would be available before long. There were queries about the degree of interaction allowed and about the aims of post-editing, but I suspect that many of us left for lunch thinking that we ought to have made better use of our chances to put some searching questions.

It was clear during the pre-lunch period, however, that searching questions were being launched at those staffing the exhibition mounted at the conference. Judging by the crowds at each stand, this was a very successful feature throughout the two days. Working equipment and brochures in plenty were displayed by Ad Tempus Computers, ITT Europe, Information Technology Limited, The Software Connection Limited, Use IT (Holdings) Limited and by ECHO (Eurodicautom) as well as by those who had taken part in the conference itself (Textnet, Logos, Weidner and ALPS). Examples of raw translation and of the software available for the production of printed documents in many different languages and formats all gave us a good feeling for the state of the art.

The afternoon began with a tour de force from Hugh McGregor Ross of Data Systems Consultants on handling non-Roman character sets with computers. His long experience in communications engineering and membership of committees on computer standards enabled him to give us a clear and entertaining account of a highly complex topic. He pointed out that character coding (i.e. the establishment of a binary coding technique for a selected character set) always meant adopting an agreed standard if there is to be any communication between one local user and another. Such encoding should if possible be on the local side of the interface with the system used for intercommunication, whether disc, tape or Telecom. Life is much simpler, of course, if the local equipment uses the same standard coding as that needed for intercom.

He then enumerated the various bodies responsible for establishing coding standards, emphasizing that two important ISO groups (TC97/SC2 dealing with general principles and TC45/SC4 dealing with bibliographical applications) lacked translators as members so that we were not influencing major developments that concerned us.

Turning to the more technical aspects, the various character codes in use and proposed were outlined: (a) the 7-bit code (ISO 646 or ASCII) allowing 12 codes for national variations, said by the speaker to be obsolete for international communication; (b) a very recent European Computer Manufacturer's Association 8-bit code, which will deal with the simpler European languages; (c) ISO 6939 with 176 individual codes and 16 control characters for work covering the more complex European languages; (d) finally, and largest of all, the Code Extension Scheme for Multilingual Use "understood by only 3 people in the world".

20

We were then treated to some examples of the difficulties encountered with various alphabets (Greek, Cyrillic, Arabic, Chinese, Japanese) just to show us the magnitude of the problems. We were convinced. The solution, we were told, was being sought in the development of a 16-bit code usable with high-level computer languages and sufficient to provide for all living languages. The talk ended with a strong plea for translators and bibliographers to take part in the development of this new code, which will satisfy all our needs only if our expert knowledge is brought to bear on it.

Character sets, however, must surely be child's play compared with the problems of voice recognition, described by Raj Gunawardana of Texas Instruments in a talk which had been supported by a demonstration the previous afternoon. Equipment being developed followed one of two main paths. First, the 'speaker-dependent' path in which the machine makes 'templates' from the voice of the user. These are stored and compared with later inputs from the same user. In principle, this type can cope with any language or vocabulary and with ordinary connected speech, but its cost would be high. The second type is 'speaker-independent' in that its templates for comparison are pre-created and stored. Such machines have a very restricted vocabulary and are often only capable of dealing with passages containing isolated words.

We moved finally to the plenary session chaired by Barbara Snell, in which a panel selected from the speakers during the two days (Birgit Rommel, Linda Talhami, Robert Rooke) were joined by Sarah Dunlop, Peter Wheeler, David Ashby and Martin Radford to answer questions submitted in advance by delegates. Once again, the Chair showed great expertise in putting the questions on our behalf, but the success of such a session involving off-the-cuff responses depends heavily on the stimulating quality of the input.

Somehow our questions did not seem the stuff of which great plenary sessions are made (or perhaps we were all just somewhat overawed by the immensity of what we were facing in the communications revolution). However, the panel were encouraging in their response to two queries about MT: it was not thought that prostitution of language would be greatly accelerated by MT, at least beyond its present rate of increase! And there was general agreement that large-scale use of MT would not lead to unemployment amongst us.

On such a happy note, Geoffrey Kingscott closed the conference by emphasizing what an important part these annual meetings play in encouraging translators to use the many aids discussed and in giving MT companies the chance to come before a demanding audience. Manufacturers of word processing systems were being less influenced by translators than they should be, however, and delegates must resolve to play their part in bringing pressure to bear on the relevant committees and companies.

BILL DUFFIN