

# **New developments in information technology for interlingual communication**

*J. C. Sager*

Department of European Studies and Modern Languages, UMIST

*Paper presented at the Aslib Technical Translation Group conference and exhibition, London 20 November 1980.*

INFORMATION TECHNOLOGY IS a new interdisciplinary field combining information science, computing, telecommunications and electronics. As it establishes links across national and linguistic boundaries it also acquires a multilingual dimension requiring translators as links in the interlingual communication process. The translator thus can no longer be considered in isolation; he is simply another mediator, generally between natural languages, similar in function to information scientists who mediate between natural and artificial languages, be they documentation languages or command languages to access databases, and similar also to those computational linguists who mediate between natural languages and computer languages. Interlingual communication is expanding and its means and methods are changing. This is not only reflected in two major Aslib Conferences on this topic in two years, but in Action Plans of the Commission of the EC, the activities of INFOTERM in Vienna, ISO, FID, and UNESCO involvement in this work, the growth of terminological data banks, as well as new organizations and studies in every developed country, some of which were mentioned during this conference.

This concern with interlingual communication is fully justified as with its increased volume new problems arise which require solutions. We all know of the cost of the multilingual regime of the European Community institutions which represents a substantial part of the budget of each organization. Equally the Council of Europe has 800 employees from 21 countries, representing 16 languages. It produces 50,000,000 written and 136,000,000 spoken words every year in any one language. With closer economic cooperation throughout the world and ever more international organizations and meetings the cost of translation and interpreting services may become too heavy a burden to bear unless solutions for reducing costs can be found. With six official languages the EC has to deal with 30 language pairs for translation, ie five translations of each significant document. With the addition of Greek the number of language pairs into and out of which translation will be required automatically rises to 42. At the beginning of this century three languages covered most publications in the field

of science and technology. A scientist with a reading knowledge of English, French and German had a nearly complete coverage of all scientific publications. While English still is the language of some 50 per cent of all scientific papers the rest is made up of some 50 other languages though admittedly of very uneven importance according to subject fields. In technology English also still has a dominant position but other language communities increasingly feel the need to develop technology in their own languages in order to stimulate home-grown research and development and to facilitate education. We are also witnessing an unprecedented development of national languages in newly independent countries which require written languages for the purpose of national development in all spheres, for some of which they had in the past relied on a European language. This is not only a national problem but requires adjustment by all other countries in regular contact with such linguistic communities who upgrade the status of their own language.

Another factor which increases the importance of interlingual communication is the growing sophistication of the export products of developed countries. In some cases the documentation accompanying a product can represent a major part of the cost and in such cases reliable translations of the full documentation is essential for the customer. Finally, the introduction of networks linking documentation databases across linguistic boundaries makes multilingual data available which, however, can be fully utilized only if translation can be provided with a speed and efficiency comparable to that of accessing the data in the first instance.

This then is the background to what is commonly described as the 'interlingual information explosion' and the various speakers have dealt with different aspects of aids for the translator. Mr Stiegler described the role and function of word processors and also looked at new developments in optical character recognition and voice analysis. It is to be noted that microprocessor applications in the monolingual office are well established. What is new is exploring their optimal use in the various processes that are globally described as 'translating'.

Mr Clark, one of the pioneers in using a word processor as translation aid, presented his own experience in using the new technology for the compilation of glossaries, interactive translation, revision, production of final copy and filing translations. He confirmed the expectations of increased efficiency and productivity achieved by the use of new tools but also pointed out that new techniques and working habits are required to maximize the use of word processors.

Miss Duckitt outlined the possibilities and advantages of online access to documentation databases. She described the services of the International Translation Centre in Delft in the logging of existing translations and how multilingual documentation databases can be used in looking for new translation equivalents. A particular kind of database, automated dictionaries or terminological data banks was described by Mr McNaught who briefly presented the scope and function of existing databases and then sketched out a model for such a new linguistic data bank as it is emerging from a feasibility study currently in progress. Such data banks will best justify their expense by being multifunctional, ie serving translators, technical writers, information scientists and standardization agencies simultaneously.

Finally Mr Arthern critically examined the prospects for the use of various machine aids in a large international organization heavily concerned with various stages of

multilingual drafts prior to publication. He also gave a survey of current efforts and plans for machine translation, computer storage of documents and, in general, the prospects for greater involvement of information technology in international interlingual communication.

A common theme in many papers was the danger of introducing new barriers into interlingual communication, barriers resulting from the requirements of the new technology itself. These barriers consist of the various documentation languages, the command languages of the various systems and different formats of data storage which prove a serious stumbling block to the free interchange of data between systems. Such difficulties have been recognized for documentation networks by the European Commission's Committee for Information and Documentation in Science and Technology which was responsible for the introduction of EURONET. The following areas of difficulty have been identified and will be investigated: the translation of the host manuals, the design of a common command language, the introduction of machine translation for titles and abstracts, a common classification system of the kind of the Broad System of Ordering and the creation of multilingual thesauri. The new directorate for the Information Market and Innovation of the European Commission has undertaken a series of relevant studies and is shortly organizing a workshop on a common character set for all Community languages. The multiplicity of tools at present available requires harmonization and standardization of components and procedures both with respect to software and the data themselves. Though millions of terms are already stored in terminological databanks they cannot readily be exchanged as there is as yet no agreement about a common data structure nor about an exchange format on magnetic tape, despite persistent ISO efforts for the last few years. There is considerable danger of unnecessary duplication of work which besides the waste of human energy will lead to conflicting terminology and translations. Agreements on a number of levels are therefore urgently required among manufacturers of hardware, designers of software, hosts of databases and producers of translations and terminology.

New information technology also requires a complete reassessment of the whole range of linguistic mediation processes and new methodologies of work. Interactive translation and text-editing are already widely used in some organizations. Split VDU screens for original, translation and dictionary display are available but manufacturers are not sufficiently responsive to translators' real needs so that there is still considerable room for improvement as more experience is gathered. Translators will also have to become more versatile to deal with pre- and post-editing of machine assisted translation and the end user will be offered a wider range of products such as raw machine output for information scanning, controlled language versions of abstracts and other relatively stereotype text forms, human revision of machine output besides the traditional full human translation. This diversification will permit a greater volume of translation to be produced more quickly but there is as yet no sizeable body of experience to ascertain the acceptability of these new forms of translation nor an analysis for which texts a particular type of translation is most appropriate.

It was my privilege to open the first Aslib symposium on this topic in 1978. The fact that I am drawing the conclusions on this occasion does not mean that the debate on this topic is closed. Quite the contrary the discussion is just beginning and

there are likely to be many more occasions like this one to monitor progress and reassess the orientations taken.

## REFERENCES

- ENGELBERT, H. The terminological barrier to the understanding of scientific and technological literature. *Translation News* 6 (2) (1976) 2-10.
- ENGELBERT, H. 'Informationsbarrieren'. *Informatik* 1974, no. 5, pp.51-54.
- GIBB, J.M., PHILLIPS, E. Scientific and technical publishing in a multilingual society, in: COMMISSION OF THE EUROPEAN COMMUNITIES, ed. *Third European Congress on Information Systems and Networks*, München 1977, 13-28.
- ILJON, A. Scientific and technical databases in a multilingual society, in: COMMISSION OF THE EUROPEAN COMMUNITIES, ed.: *Third European Congress on Information Systems and Networks*, München 1977, 28-33.
- KUNICKI, M. La barrière linguistique, son importance et son évaluation. *Documentaliste* 17 (1980) 4-5, pp.147-150.
- SAGER, J. C. and others. *English special languages—principles and practice in science and technology*, Wiesbaden: Brandstetter Verlag, 1980.
- SAGER, J. C. The Computer and Multilingualism at the European Commission. *Lebende Sprachen*, 3, 24, 103-108.
- TYBULEWICZ, A. and LIEBESNY, F. The relative importance of various languages in scientific and technical literature. *The Incorporated Linguist* 4 (1965) 1, 12-13.
- VAN BERGELJK. D. and RISSEUW, M. The International Translation Centre, The language barrier in the dissemination of scientific information and the role of the ITC. *Journal of Information Science*, 2 (1960) pp. 37-42.
- VAN SLYPE, G. *Les langages documentaires dans la perspective de l'utilisateur des réseaux internationaux d'IDST*. Paper presented at the 39th FID Congress 25-28 September 1978, Edinburgh, Scotland. London: Aslib, 1980.
- WELLISCH, H. Linguistic and semantic problems in the use of English language information services in non-English-speaking countries or how to install an elevator in the tower of Babel. *International Library Review* 1973, no. 2, p. 147-162.
- WOOD, D. N. The foreign language problem facing scientists and technologists in the United Kingdom—report of a recent survey. *Journal of Documentation*, 1967, no. 2, p. 117-139.