

## The Linguist as Technologist

by

John McNaught,  
Centre for Computational Linguistics,  
UMIST, Manchester, UK.

The argument of this paper is based on two fundamental assumptions which my colleagues and I view as incontrovertible.

These are

- 1) it is desirable to translate non-literary texts
- 2) it is desirable to train linguists to participate fully in the translation process.

In this paper I would like to address myself primarily to those educators who are responsible for the formation of linguists. I should like to consider the role of the linguist and offer a re-appraisal of that role in the light of recent advances in communication technology.

### AUTOMATION IS A REALITY

Over the last decade, dramatic advances in technology have revolutionised the transmission and handling of information. The most characteristically "human" aspect of information - and consequently the most difficult to mechanise - is language, and particularly the conversion of information from one language to other. Linguistic sensitivity and elegance of expression are qualities which make the work of the first-rate linguist as much an art as a skill, and it is unthinkable that these qualities should ever be supplanted by a machine. Yet automation has begun to affect even this most human of activities. There are pressing reasons for this desire to mechanise certain aspects of the communication process. These may be summed up in two words: "information explosion".

Generally, on a world scale, there exists a considerable backlog of texts requiring translation. To take the particular case of the C.E.C., where every document of general application

must be produced in six languages, very short deadlines have a detrimental effect on the style and quality of translations (Arthern 1978, p. 81). This, in spite of the fact that a substantial part of the Institutions' budget is for linguistic services (Sager 1978, p. 5), and approximately 1700 translators are employed full- or part-time (Vauquois, 1979). Furthermore, Anderla has suggested that by 1987 the world's annual production of scientific and technical information will be six times the then current rate (Anderla, 1973).

It is a simple fact that the output of such information is beyond the physical capabilities of human translators. Even if the necessary numbers of human translators were available, it would prove impossible to employ them, for a variety of reasons:- considerable expense would be entailed; space would have to be found for these people to work in; severe problems of organisation would arise; consistency of translation in large projects could not be guaranteed; more paperwork than ever would be generated; the pressures of working in such conditions would be great.

Thus, automation is seen as the only answer to this otherwise intractable problem. Automation of the translation process has been in progress now for several years. The computer has become an indispensable tool - it is here to stay, for there exists no other viable alternative. All linguists, and especially those we are training at present, must learn to accept and to use the new aids technology will provide. That automation is a reality cannot be too strongly emphasized. Refusal to accept a degree of automation in his sphere of activity can only do the linguist a disservice in the long term, for the machine is here to stay - it will not go away. Refusal to come to terms with the computer will result in systems continuing to be designed by people, who, however well-meaning, have no appreciation of the problems or needs of translators; thus it is in the translator's own interests to explore the possibilities offered by the machine, and to make sure he can exploit these to his full benefit.

#### APPLICATIONS

Let us explore then, some of these computer aids for translators.

I shall mention only a few fields where machines are actually doing vital work of immediate relevance to the translator.

*Machine Translation* (M.T.) is, perhaps, the most glamorous and best-known application of computational linguistics. I shall however refrain from going into M.T. in detail, as several papers will be presented on this subject during the course of this conference. (See the papers by R.L. Johnson, G.C. Keil, H. Ruus and B. Maegaard.)

Continuing developments in long-distance information communication systems will also affect the translator. National post offices and T.V. companies are setting up information networks which will bring a vast amount of knowledge into one's own home or office. The C.E.C. is currently implementing plans to establish a European information network for scientific and technical information (EURONET), which will link data bases in several member countries, and will be accessed via post office communications. One of the first services to be made available to EURONET users will be the C.E.C.'s own Terminological Data Bank (EURODICAUTOM).

Mention of data banks introduces an area which is perhaps more familiar to the professional linguist - the storing and accessing of terminological information by computer. *Term banks* provide a wide range of different services to many kinds of user: translator, terminologist, specialist librarian, technical writer, language teacher, abstracter, lexicographer, etc. They came about due to a desire not just to store information for its own sake, but to store information on technical terms, industry standards and the like, and to provide means of accessing these data in a variety of illuminating and rapid ways. This automatically implies that term banks are not tailor-made for the translator, and indeed some dissatisfaction over services provided by term banks has been expressed. Several criticisms may be found in Rondeau et al (1977) (rather damning) and Arthern (1978) (pros and cons).

Nevertheless, the number of term banks is increasing constantly, and the quality of their services is such as to allay criticism. For comprehensive listings of established and proposed term banks and accounts of their services, see Carestia-Greenfield & Serain

(1977), Sager (1978), Sager & Johnson (1978). Term banks catering specifically for the translator include EURODICAUTOM (Goetschalckx, 1978), Bundessprachenamt (Krollmann, 1977), Siemens AG (Tanke, 1978), Montréal (Dubuc, 1975, Paré, 1974) and Québec (Fortin, 1974). In Britain, UMIST has recently been awarded a grant from the British Library to conduct a feasibility study into the establishment of a British Term Bank. In Sweden, TNC in Stockholm have set up their own term bank, TERMDOK. Here in Denmark, CEBAL are in the process of establishing DANTERM, a description of which may be found in the papers submitted by Lene Frandsen and Bodil Nistrup.

Is there however a real need for term banks from the translator's point of view? Let us look at some of the consequences of *not* using a term bank, and let us take a simple everyday task: that of discovering the appropriate translation of a new term. The translator's trusty companion, the bilingual dictionary, proves to be of little help in this instance, since, when he bought it, it was probably already out of date. The translator may have to wait several years before the next edition, to find the required term - of scant use when the translation is required immediately! And moreover, by the time the term does appear in print, a revolution in the technology of the appropriate field may well have made the term obsolete. The translator may then have to resort to other methods: consulting other translators' files, or asking an expert in the field. This may prove no help, or produce several different translations. A further method would be to look for the term in the specialised vocabulary of a different field, with the danger of procuring a misleading or totally false rendering. As a last resort, the translator may coin a neologism. (Coining of neologisms may well have been the reason for several renderings being found in other translators' files.) Several translations for a single term together with the proliferation of neologisms raises the problem of *consistency* of translation.

Further, it has been found that a translator spends as much as 60% of his time searching for terminological information in published dictionaries (Lippmann, 1975). What is more worrying, in view of this inordinate amount of valuable time devoted to research, is that according to one study (Krollmann, et al, 1965), 62.1% of translation errors were found to be caused by terminological

inaccuracies. With the information explosion already upon us, and the ever-increasing demand for translations, it is clearly in the interests of all participants in the translation process to ensure that translations can be produced quickly, efficiently and accurately.

I would argue that this may be achieved through use of a term bank. Consider some of the services provided by a term bank and the advantages it represents over the use of conventional published dictionaries. A properly constituted term bank should provide the translator with, firstly, a rapid means of accessing a far wider and potentially more informative range of terminological data than that contained in a published dictionary (which has the added disadvantages of being bulky, time-consuming to use, soon obsolete, prone to wear and tear, and not particularly cheap); and secondly, various means of presenting data to suit special needs.

The ideal term bank will be equipped with computer dictionaries and accessing systems that are organised in such a fashion that the translator receives just the right type and amount of information that he wants. (Criticisms levelled at early term banks were that they provided either too much or too little information, which was distracting and misleading.) A well-defined *interactive* interrogation system provides the translator with several options. It should allow him to pursue a term's meaning and translation, and the relationships it contracts with other terms, by permitting him to widen or narrow the search space. It should provide him with contexts, definitions, sample usages of the term, or a display of the term in a thesaurus-type hierarchy with related terms. It should also provide synonyms, antonyms, and near equivalents, indications of style and register associated with the term, and statistics on, for example, the frequency of the term.

The systems of the Bundessprachenamt and of Siemens, for example, provide useful means of organising information. At the Bundessprachenamt, the user may request among other services the following: firstly, printout of a text-oriented glossary, which provides translations for specified terms in the order they occur in the text. A study into the use of such an aid found that

translators spent a third less time on a translation and made only two-thirds of the errors registered by translators working conventionally (ALPAC, 1966). Secondly, alphabetic printout of a text-oriented glossary is useful for ensuring terminological consistency when work is split amongst several translators. Thirdly, microfiche of any part of the term bank is provided. These are cheap and easy to produce and are ideal for short translation jobs. They may be replaced at regular intervals, which will ensure that the translator has a fairly up-to-date record of the term bank's holdings in his particular field. About 8000 terms can be stored on a microfiche of 10 x 15 cms. For a detailed account of the services offered by the Bundessprachenamt, see Krollmann (1977).

The Siemens system likewise produces text-oriented glossaries and various alphabetical lists, on microfiche or hard copy. Moreover, it provides an ideal example of how a term bank may centralise information, yet make it widely available, for a Co-operative has been formed, which includes Siemens staff, freelance translators, private bureaux, etc. Terminology is stored by Siemens and is made available through various communications systems to other members, who in return play a vital part in contributing data to the bank. For further information regarding this system, see Tanke (1978).

The impact of the *microprocessor* has been such that today a freelance translator may aspire to possessing his own computer. Within a few years, the home computer will be commonplace. For the translator, this means that he will have access to the same kind of services at present provided by term banks, though on a smaller scale. Nevertheless, substantial mono- / bi-lingual dictionaries may be bought on a series of *floppy discs*, at a fraction of the cost of a conventional published dictionary. (A floppy disc is akin in function to magnetic tape, that is, it is a storage medium. Between 5 000 and 10 000 terms may be stored on a floppy disc, depending on the compression techniques employed.) Such dictionaries could of course be provided by a term bank, and replaced with up-dated versions at regular intervals... Micro-computer prices are decreasing rapidly, to the extent that the home computer is a definite reality. With his own system, the translator will of course be able to add or delete terms from his

dictionaries at will. Note that access time is usually faster on a small computer with only one terminal than on a mainframe computer run on a time-sharing basis. Moreover, there exist already hand-held computers that provide a limited number of translations. While these are still very much at the novelty stage (Blumenthal, 1979), their development should be followed closely. Today's toy has a habit of becoming tomorrow's tool.

The computer is proving of immense help in organising, storing and disseminating information in other fields. Computers are commonly used now to facilitate document retrieval. Question Answering systems enable interrogation of large data bases in natural language. Automatic systems are used to provide abstracts of journals. Machines are to be found now wherever language is. They range from the humble word-processor to the "intelligent" computer, which is capable of 'understanding' natural language instructions and simulating processes and activities thought to be the prerogative of humans.

#### EDUCATION IS A NECESSITY

The linguist, therefore, must learn to live with the machine, he must learn how to use the machine. Those linguists we are actually training will have to feel at home with the computer, for within a few years, to take but one example, such a commonplace as the published dictionary will in all probability become a rarity.

All this implies that *teachers* will have to come to terms with these new methods, and that quickly. They will have to cultivate a different outlook, a different approach to the whole field of communication and translation. If we continue to go on as at present, then we cannot but fail to achieve proper communication in the future. In particular, we must ensure that a certain *information barrier* is overcome, that between computer scientist and linguist.

Let me illustrate the problem with an analogy: that of the preparation of a dictionary, which involves, let us say, someone who knows about language and who wishes a reference tool - the linguist - and someone who knows how to overcome the technical

problems involved - the printer. It would be ridiculous, I am sure we would all agree, if the linguist were to ask the printer to collect data, and organise these in linguistically relevant ways. The printer has no training in such a task. Any dictionary he produces according to this method will inevitably be unsatisfactory to the linguist.

Yet until very recently this was exactly how computer systems for language handling were designed. Indeed, systems are still being designed by people who have little or no linguistic training. A computer scientist may know his subject intimately, but he is rarely competent in linguistic matters. Linguists asked to use such a language handling system designed wholly or primarily by non-linguists often feel alienated - for the system does not and cannot give the linguist the information he needs to do his job. Being compelled to use such a system by an employer who has invested a large amount of money in it, results in the linguist producing poor quality work. He will also develop a resentful attitude towards the system, which will prejudice him against all machines and all systems, no matter how good.

It is true that many of the tools we have at present are deficient. The defaults of, for example, EURODICAUTOM are well-known, one of the main complaints being that this system does not provide the right type or amount of information sought by the linguist. This system was in fact designed by data base specialists and terminologists. Carestia-Greenfield & Serain (1977) and Arthern (1978), among others, both note that there is an unfortunate resistance among translators to computer aids, and especially to interactive interrogation. This is understandable if the objection is that the desired information is not supplied. However, resistance to the computer for subjective reasons must be overcome, for even at this moment difficulty would be experienced if computer aids to translation were not available. Krollmann states that "by and large it is true to say that despite initial opposition on the part of the translators, who at first were sceptical towards the computer, the day to day work of our translation service would now be *inconceivable* without the system" (1977, p. 245; my emphasis). Arthern (1978, p. 93) makes a plea for what he calls a "translator package" which would provide on the screen of a Visual Display



Unit the minimum amount of information usually required by translators. This plea arose from his experience of EURODICAUTOM. Carestia-Greenfield & Serain (1977, p. 93) likewise suggest that user resistance to computers would be overcome if systems were designed with the end-users' needs in mind.

#### THE LINGUIST AS TECHNOLOGIST

What is lacking, to all intents, is an *intermediary* - someone who is as much at home writing programs as writing grammars, someone who understands the problems of the computer scientist, of the terminologist and those of the linguist.

Such a person fills a crucial role - there are as yet few designers who have both the necessary deep understanding of linguistic problems and the technical comprehension of the capabilities of the computer. It is our task as educators to

- come to terms with machine handling of language
- realise the need for a *new breed of linguist*
- ensure that our courses provide training and encouragement for this important and highly specialised linguistic work.

We shall be training an elite class - of people who will make a vital contribution to the information handling process. This role is indeed crucial, for if it is not filled, then the computer scientist and the terminologist will remain divorced from the translator. This means that tools will be provided which will continue to alienate the user, in that he will not be able to perform the operation he wants; he will only be able to complain bitterly about lack of help - he will *not* be able to talk to the computer scientist on his own terms and work out a solution.

There is no shame whatsoever attached to becoming a member of this new linguistic elite, to becoming a 'technologist'. Every good technologist is at heart an artist. He is a creator. He exhibits a certain flair in his work. He takes pride in his work. At the end of the day, he produces something of direct use, something that in our case is in increasing demand. In his practicality however, there is elegance, there is a concern to

provide a habitable interface between man and machine, to put machines at the service of man.

The linguist of the future will have to develop, as Sager puts it, "new attitudes and working habits" (Sager, 1979, p. 105), which will mean a reappraisal of present courses. Johnson has succinctly summed up the qualities required by the linguists of the future:

"The new linguistic elite will need to know one or two foreign languages, as well as their native language, both dispassionately as objects of study and intimately as means of communication, to handle the symbols of language with the ease and objectivity with which a mathematician handles the symbols of an algebra, and to be at least a competent computer programmer. Such a training has little regard to the traditional, artificial distinction between 'Arts' and 'Sciences' which still pervades educational systems. Like a good engineer, the new linguist will be creative in his design and rigorous in its execution, working within and not against the progress of technology."  
(Johnson, 1979)

The role of the new breed of linguist in fact goes far beyond simply the design and implementation of aids for translators. A substantial proportion of all the information stored, processed and transmitted by computer is originally expressed in language, yet paradoxically the implementors of information systems rarely include trained linguists with a professional appreciation of the complexity of the data which are being processed. Translation by machine is only one small corner of a vast field of linguistically oriented technology, whose success will depend crucially on the active and informed participation of the next generation of linguists. You will hear tomorrow about M.T., which is already beginning to affect the role of translators. I do not wish to elaborate on this; I should merely like to point out that, in our view, the biggest single reason for the ill-starred fortunes of M.T. was exactly this barrier between computer scientist and linguist, the lack of an intermediary, the lack of a linguist who is a technologist.

## REFERENCES

- ALPAC (1966) *Languages and machines: computers in translation and linguistics. Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Publication 1416, Washington DC.*
- Anderla, J.G. *Information in 1985: a forecasting study of information needs and resources, Paris (1973):* OECD.
- Arthern, P.J. (1978) 'Machine translation and computerized terminology systems - a translator's viewpoint', in Snell (1979), 77-108.
- Blumenthal, R. (1979) 'Testing out two pocket electronic translators', *New York Times*, March 25, 1979.
- Carestia-Greenfield, C. & Serain, D. (1977) *La traduction assistée par ordinateur: des banques de terminologie aux systèmes interactifs de traduction, Paris-La Défense/le Chesnay: AFTERM/IRIA.*
- C.E.C. (1977) *Overcoming the language barrier, München, Verlag Dokumentation, 2 vols.*
- Dubuc, R. (1975) *Description du système TERMIUM, Montréal: Université de Montréal Banque de Terminologie.*
- Fortin, J.M. (1974) *La banque de terminologie du Québec, Québec: Banque de Terminologie.*
- Goetschalckx, J. (1977) 'Terminological activities in the European Institutions, with special reference to EURODICAUTOM', in C.E.C. (1977), vol. 1, 123-152.
- Johnson, R.L. (1979) *Multilingual Communication and the Technology Revolution in Applied Language Studies & Multi-*

*lingual Communication*, UMIST. Proceedings of a conference, internal document.

- Krollmann, F. (1977) 'User aspects of an automatic aid to translation as employed in a large translation service' in C.E.C. (1977), vol. 1, 243-257.
- Krollmann, F., Schuck, H. and Winkler, U. (1965) 'Herstellung textbezogener Fachwortlisten mit einem Digitalrechner: ein Verfahren der automatischen Übersetzungshilfe', *Beiträge zur Sprachkunde und Informationsverarbeitung*, No. 5, 7-30.
- Lippmann, E.O. (1975) *On-line generation of terminological digests in language translation: an aid in terminology processing*, IBM Research Communication, September 1975.
- Pare, M. (1974) *The University of Montreal's Banque de Terminologie*, Nice: International Federation of Translators.
- Rondeau, G., Grégoire, J.-F. & Tessier, P. (1977) 'Banques de terminologies et linguistique computationnelle', *META*, 22, 184-191.
- Sager, J.C. (1978) 'Multilingual communication: Chairman's introductory review of translating and the computer', in Snell (1979), 1-25.
- Sager, J.C. (1979) 'The Computer and multilingualism at the European Commission', *Lebende Sprachen*, 3, 103-107.
- Sager, J.C. and Johnson, R.L. (1978) 'Terminology: the state of the art', *AILA Bulletin*, 22, 1-12.
- Snell, B.M. (ed.) (1979) *Translating and the computer*, (proceedings of a seminar, London, 14th November, 1978),

Amsterdam, North-Holland.

- Tanke, E.H.  
(1978) 'Implementing machine aids to translation',  
in Snell (1979), 45-69.
- Vauquois, B.  
(1979) *Aspects of mechanical translation in 1979*,  
(paper for IBM Scientific program, Japan, July  
1979), Grenoble, GETA.