Surrey University terminology project

by Margaret Rogers and Patricia Thomas

KITES = Knowledge Based Integrated Terminology System

In the November 1987 issue of Language Monthly Alex Gross, chairman of the machine translation committee of the New York Circle of Translators, wrote a challenging letter in connection with electronic glossaries, which are also known as terminological data banks or term banks. Two researchers based at the University of Surrey reply to his challenge, and describe the work of the KITES project with which they are involved.

Alex Gross maintains that ambitious claims are being made for the immediate future of "electronic glossaries". It is not clear who is making these claims: certainly not the people who are involved in the day-to-day development of such systems. They are only too aware of the problems and would never pretend that the "equation" (Mr Gross's metaphor) has finally been solved. No "solution" exists which is perfect, since the development of the terminology in term banks is a dynamic process.

At the moment, considerable work is being invested in the development of computational tools and techniques, and linguistic models for the representation of terms and of the knowledge of specialised subject domains to which these terms refer.

Furthermore, subject domains themselves are continually developing (the dynamic process mentioned above). These two factors together — development of the term bank "shell", and the changing nature of subject domains — mean that research and development will continue to be necessary, even if the balance of emphasis may shift in the future towards accommodating changes in subject domains. In fact, Alex Gross himself later in his letter makes the observation that the development of "electronic glossaries" may be rather like painting the Brooklyn Bridge. However, we do not accept the negative connotations of this analogy. Would Mr Gross prefer that Robert Burchfield had never started on his supplements to the Oxford English Dictionary (1957–1986)? Or even that Sir James Murray had not embarked on the original New English Dictionary on Historical Principles (1888–1933)?

Mr Gross illustrates many of his points by reference to a putative bilingual German–English, English–German glossary for "all the sciences and social studies". This he calls a "simple" example. Presumably he derives his notion of "simplicity" from the fact that this putative system is bilingual and not multilingual and that the two languages concerned are closely related and have both played an important role in the development of these subject domains. But Mr Gross is surely teasing us! Is it possible that he really thinks that the development of fully researched glossaries in all (!) the sciences is "simple", even monolingually? We doubt it. Also, as any translator knows, non-taxonomic subject areas (such as his "social studies" example) are notoriously difficult to systematise. Is Mr Gross's purposes to flush out a confession here that such a task is difficult? If so, he now has it. But serious researchers do not pretend anything else.

Mr Gross's point about consistency of updating is one which needs to be taken seriously. Implemented systems need to be protected from anonymous and unauthorised updatings. This is not difficult provided the problem is recognised in advance. The importance of liaison between terminologists and subject experts needs to be stressed and considerable time needs to be devoted to it. Mr Gross mentions this, but seems to regard it as an exceptional case. In the development of a reliable term bank it is essential. Furthermore, in a conceptually-based term bank, i.e. one which is based on terminological rather than word-based lexicographical principles, gaps and overlaps are more easily avoided or at least highlighted.

Mr Gross is concerned that terms used across domains (he is thinking perhaps of examples such as *nucleus*) are a problem. Certainly, the differences can be flagged, although there is a problem, when the meaning from one domain is used in another domain. Mr Gross seems concerned that the translator/expert may have to solve this particular problem. But why not use the human intelligence of the user? We have to recognise that our priorities in system development and implementation must be guided by cost-effective considerations. We are realists after all.

Mr Gross seems to want us to realise that his "simple" example of a glossary specification is in fact a complex one. He is right. He emphasises his point by indicating the even greater difficulties of designing and developing systems for less well matched languages and cultural systems. He even alludes to "omnidirectional" glossaries, by which we understand him to mean multilingual glossaries. The answer at present is simple (sic): concentrate on bilingual glossaries in defined subject domains. This may mean a user is involved in more looking up compared with fully integrated global systems, but surely this leads to greater accuracy.

We are not quite certain specifically when Mr Gross is referring to in his remarks about "universal grammar" and "deep structure". Certainly these concepts have been important in developing machine translation systems, but their role in the development of term banks is unclear to say the least. Perhaps Mr Gross is thinking of various semantic theories which may be of relevance in developing relationships between terms. One area of development where we envisage various aspects of linguistic theory (e.g. functional sentence perspective, text linguistics) to be of potential benefit is in intelligent retrieval of information embedded in the definition (i.e. a text) and this we are aiming to incorporate in our project.

In Great Britain, part of the University of Surrey KITES Project is involved in building such "electronic glossaries". We have examples from five subject domains Virology (EN/FR), Human Computer Interface (DE/EN), Automobile Manufacturing (Four-wheel drive) (DE/EN), Local Government Administration (EN/NO), Sound Insulation (EN/SP/DE/FR) which have been treated bilingually, and would welcome high quality

terms which have been verified by leading experts. If such terms could be made available to us, we can supply a comprehensive term record format with guide-lines for its completion, which we can link to the MATER format. There is a pressing need from translators outside the UK for high quality British English equivalents.

In addition to the electronics glossaries mentioned and the elicitation of the required information only, KITES aims at forming an important part of an integrated office system in a European environment.

Mr Gross also raises a number of more practical problems. Economic arrangements are indeed a problem. Floppy discs with updates could be sold or subscribed to at regular intervals. As mentioned, maybe a global on-line rental system could be operated.

The idea of compatible standards in the field of computing will be welcomed with open arms, we feel sure — and to avoid idiosyncracies, standardised glossaries are required. Errors must be the responsibility of the compilers who should be named and to whom recourse can be made.

An international clearing house for terminology (an extension of the considerable work done by INFOTERM in Vienna, for example) might provide one solution. On a national level, a state-run system such as that found in Eastern European countries produces standardised terminology in at least that country's language.

Translators will understandably be reluctant to donate their own laboriously constructed glossaries and payment directly for these is likely to be sparse. The answer is to employ terminologists who will do the background research work for individuals and who are at the same time responsible to a term bank where the new terms can be inserted. The question of payment of royalties remains a tricky one and the solution in the last sentence may avoid this.

Finally, terminology as "the heritage of all human culture" is a little too philanthropic to stomach; terminology is a commodity, not an ideal and as such demands adequate remuneration.

© January 1988, Margaret Rogers, BA, MB, Patricia Thomas, M.Phil, MIL