Machine translation: Getting personal with LIDIA

"MT will have to get interactive." We've heard that for some time now, without it ever being specified whether this meant upgraded rigs modelled on the interactive ALPS or MicroCAT software, or the latest manifestation of multimedia.

News from the research front suggests that interactivity is a new paradigm for handling translation. This is gradually being implemented as the "tough guy" approach to MT – big commercial systems crunching out rapid copy on an industrial scale for post-editing – hits the knowledge barrier.

But to guarantee powerful disambiguation routines for these mastodons, developers are discovering that the programs need more and more knowledge – linguistic, encyclopedic and pragmatic.

Japan's fifth generation computer project, for example, is committed to producing massive "knowledge bases" as a necessary resource for advanced information processing using natural language understanders. The trouble is, no one is sure of the right underlying architecture for a knowledge base, and even when they opt for one, developing a major subset of knowledge is going to cost a small fortune or two.

Forget it, says Christian Boitet, "Personal MT" is turning the tables on 35 years of this approach. Down in Grenoble, France, Boitet's experienced GETA translation team has a project aimed at harnessing the revolutionary resource of a hemoglobin-driven knowledge base to help the computer do its job. The project aims to "democratize" MT using people power.

Boitet's project, called LIDIA (Large Internationalization of Documents by Interacting with their Authors), aims to develop a "consumer" MT product for writers who don't know a word of the system's target language.

Boitet says the best knowledge base for NLP is inside our heads – so we should use a computer to guide a human, using his or her knowledge as an aid to writing and translating, rather than trying to cram the whole works into an autonomous MT system.

"We've had techno-scanning systems (Systran used by the US Air Force) and currently MT for the post-editor," explains Boitet. "Now we want to harness human knowledge together with MT processing power in a PC environment for the rapid transfer of text where no post-editing is possible."

FUTURE MARKETS

If this sounds utopian, remember that the forerunners of this approach include the two English-to-Japanese projects developed by Sheffield and UMIST (Manchester) Universities. They were part of the UK's Alvey IT program during the mid-80s under the leadership of George Jellinek and Peter Whitelock.

Developed for a project known as *Read and Write Japanese Without Knowing It*, AIDTRANS and NTRANS (the two prototypes) offered such MT resources as alternative semantic readings of sentences (and not just words). This choice of readings allowed intelligent human users to make a selection according to their understanding of the whole document. It is not yet known whether Sharp, who put up some of the money for the project, will take these rigs into the development stage.

The GETA approach is better geared to technology trends than the TRANS rigs, however, since it uses the HyperCard environment on Macs to develop an interface for the human writer when producing, say, a technical documentation manual complete with graphics.

"HyperCard is ideal for this process," explains Boitet, "because as a set of linked objects it lends itself to interactivity. Trying to develop a writing station on the classic wordprocessor would be to misread the technology."

The enhanced interactive system will be organized around a dialogue with the writer who will be prompted to "standardize" input text and provide clarifying responses to machine requests to disambiguate text material lacking adequate syntactic, semantic or discourse structural clues.

This writing interface on the Mac will be linked to a multitask system with a batch spell checker and other text-critiquing resources. The whole system is linked online to a MT server once the input text has been cajoled into shape by the combined talents of writer and computer. The MT system used to grind out the actual translation will be one of the rigs developed over the years by the GETA team.

The LIDIA project also aims to integrate a high-quality speech synthesis module so that the machine can guide the writer in real time via spoken messages, as well as providing a synthesized output to the source language text as an auxiliary spell and style checker.

The system will also include a reverse translation module: once the text has been made fully machine readable by the combination of human and machine software, it can be translated and then "back-translated" to offer the user a quality assurance facility. With an eye on future markets perhaps, GETA is planning to offer French into Russian and German. After checking the back translation, the final document can be sent over the modem to that remote printer in Murmansk or Dresden for the ultimate end-user.

If LIDIA stays the course, we might find one day that Personal MT spells the move from CAT to HAT – Human Assisted Translation.