

Translator Workbench

by
Geoffrey
Kingscott

The *Translator's Workbench* (TWB), a multi-operation work-station created by a consortium of European industrial companies and universities, was unveiled at a three-day seminar held in Heidelberg, Germany, in November.

The TWB project, supported by European Community funds under the ESPRIT programme (it is Project EP 2315), involves multilingual text processing, terminology database creation and interrogation, and machine-aided translation, all brought together and accessible from the same screen and keyboard.

The organisations involved in the consortium are TA Triumph-Adler AG, the equipment manufacturer, of Nuremberg, who act as the prime contractor; the Fraunhofer Gesellschaft, a research institution, of Stuttgart; L-Cube Information Systems SA, a software firm, of Athens; the German and Spanish data processing company of Siemens-Nixdorf, in Munich and Barcelona; the universities of Heidelberg and Surrey; and the Polytechnic University of Catalonia.

The participants, experts in translation and terminology from a number of European countries, were welcomed by *Peter Hellwig*, professor of computational linguistics at Heidelberg University.

Gerhard Hever, of Triumph Adler, said that the project had begun in April 1989 and was due to be completed in March 1992. The work station's facilities would include multilingual text processing, the capacity to process documents in the ODA and ODIF formats, the use of various advanced language checkers for pre- and post-editing, terminology management, remote access, and some automatic pre-translation.

Guest speaker *John McNaught*, of the Centre for Computational Linguistics at the University of Manchester Institute for Science and Technology, in what he announced as a deliberately provocative paper, criticised the lack of practical relevance in much university computational linguistics (CL) research. He did, however, acknowledge that the understanding of natural language processing (NLP) problems had grown tremendously in recent years, in particular the problem of the interaction between linguistic knowledge and extra-linguistic knowledge.

He passed in review various aspects of CL, and commented on the progress or lack of progress being made towards relevant applications. Mainstream NLP research, he argued, was not yet capable of delivering prototypes that can tackle 'real language' with any great degree of success:

Current NLP system design strategy relies on evolving descriptions that are built up largely to respond to some notion of general language. Systems that then appeal to such a notion demonstrate a marked lack in quality when called upon to process "real language", particularly texts dealing with technical and scientific subjects....The bulk of the work that has been undertaken in NLP research, from the linguistic description point of view, is of little relevance to being able to process the types of text that NLP systems are typically called upon to tackle. We appear to have good tools, in the form of formalisms; however, there appears to be a lack of adequate linguistic knowledge to be manipulated by such formalisms.

He called for the main thrust of effort to be made in the research area of linguistic descriptions, and in addressing the problem of tackling real language in all the varieties found in texts of interest to NLP. He considered that one hopeful direction was the work carried out by small groups on sublanguages and corpus processing.

On the second day there were both demonstrations of the TWB in action, and talks about different aspects of the system. *Monika Höge*, of Mercedes-Benz, described how users had become involved in the development of TWB, which had throughout taken a user-driven approach. Mercedes Benz and the University of Surrey, for example, had collaborated in a survey among professional translators in Europe to assess how they could benefit from advances in information technology. *Renate Mayer*, of IAT Stuttgart, described how the TWB user interface had been designed to be as ergonomic as possible. *Gregor Thurmair*, of Siemens-Nixdorf, surveyed research in spelling and grammar checkers, and verification of controlled language, carried out in the various centres for the project, while *Peter Hellwig* described an experimental parsing and processing system, based on Dependency Unification Grammar, being developed at the University of Heidelberg, known as PLAIN+ (Programs for Language Analysis and Inference), and which has been incorporated into the TWB.

Jaime Delgade, of the Polytechnic University of Catalonia, dealt with TWB access to external resources, such as the terminology database EURODICAUTOM, and MT systems, such as METAL, via X.400, using ODA/ODIDF as the document interchange format.

On the third day, in a session chaired by *Khai Le-Hong*, of Mercedes-Benz, *Norbert Brinkhoff-Button*, a member of the Commission of the European Community Eurotra team in Luxembourg, spoke

about the EC's new LRE (Linguistic Research and Engineering) initiative, whose goals, possibly reflecting some of John McNaught's concerns, are to promote a shift from theoretical computational linguistics to more practical linguistic engineering. The initiative has a budget from projects of 20 million ECUs over a period of three years. Its administration will be similar to that of the successful ESPRIT programme.

A final panel session looked at possible future developments in multilingual text processing. *Ulrich Heid*, of the University of Stuttgart, emphasised the

need for "multifunctionality" of tools for language professionals. *Cornelius van der Horst*, of the European Commission, Luxembourg, pointed out that there was little in between "the bulldozer of MT and the expensive human translator". He suggested one possible promising direction would be to make source texts more easily translatable. The translator or MT post-editor, he said, "is like the man who sees a baby in the water, and dives in and rescues it. Then he sees another baby, and rescues that; and so on and so on. No-one ever thinks of going upstream to stop whoever it is who is throwing the babies in the water".