

True European Translation Tools

Participants and observers believe the Translator's Workbench Project may be one of the most successful ESPRIT programs to date. Will the market agree?

"Computer-aided translation tools are progressing, but they're not there yet. They're not yet answering users' needs in real-world applications," said Triumph-Adler's Gerhard Heyer, at the opening of the Translator's Workbench Workshop (TWB) in Heidelberg last November, where the members of this international consortium gathered to discuss their efforts and demonstrate prototype systems. Software developers, he explained, such as his group at Triumph-Adler Research, are now directing their efforts towards wordprocessing enhancements because the market for wordprocessing packages is nearly saturated. "Any such extensions," he warned, however, "must be based on existing software, especially in the PC world." Heyer went on to point out that the unified European market could mean a huge market for writing tools aimed at people writing in a second language. Some tools are now available, such as the American grammar checkers, but they do not go nearly far enough in this direction, he believes.

The TWB project, which officially concludes March 31, represents a noteworthy new direction in Community-funded programs. As Toon Lowette, an independent management consultant in Brussels who evaluates the commercial feasibility of such projects put it, "one of the criticisms of ESPRIT in the past has been the lack of concrete results. There's been too much emphasis on the R-for-research. TWB is a welcome departure from that." The project is decidedly product-oriented and has an innovative structure, whereby end-user input has been integrated from day one in the design and implementation of the prototype. TWB participants have developed a suite of software modules to agreed-upon specifications yet maintain full, motivation-stimulating commercial rights to their efforts. As befits an ESPRIT project, TWB appears to have been a fruitful union between industry and academia, from both north and south of the Alps. Less characteristic for an ESPRIT project, however, yet all the more welcome, is that "competitive" products will directly result.

TA's multilingual editor

the core of the system is a multilingual editor, developed by prime project contractor Triumph-Adler, initially in Framemaker under UNIX, but now also in Word for Windows, using the programming language of that package, WordBasic. The editor offers such translator-friendly niceties as source and target text windows with parallel scrolling and multilingual proofing facilities. Discussing the editor, Triumph-Adler's Marianne Kugler reinforced the increasingly more accepted argument that you have to address translators' needs on their home turf; it costs them too much time and effort to learn a new editor. Moreover, it is no longer considered acceptable to sacrifice formatting. On its own, Word for Windows (whose DOS-based predecessor has about a ninety percent share of the WP market in Germany) may not increase productivity, Kugler explained. Rather it is the online facilities, with cut-and-paste between programs, and import/export options which are the key advantages of an integrated multilingual text processing and translation environment.

Existing and proposed products which could then be used in conjunction with the editor include Siemens-Nixdorf's TermPC and Surrey's MATE for terminology management and proofreading modules for Greek, German, and Spanish being developed by L-Cube, Triumph-Adler, and Siemens-Nixdorf. Spellcheckers are fairly common nowadays, pointed out SNI's Gregor Thurmair, so TWB tries to offer more, both laterally, in terms of language coverage, such as Greek (which Thurmair describes as "exceedingly difficult with its rich inflectional behavior and many irregularities"), and vertically, in what he calls "recognition and correction strategies." The extended German spellchecker developed by Triumph-Adler checks irregular compound words and capitalization; it also checks agreement within noun phrases by means of an ATN-based parser.

Additional functionality is offered by Fraunhofer's flavor of the Translation Memory concept, which in this implementation uses statistical data acquired in a user database to recover words, phrases, and sentences which have occurred in previous translations. To ensure that TWB users can communicate with the external world, a group at the University of Catalan has developed format conversion software and designed telecommunications links to provide users with remote access to METAL and EURODICAUTOM. As the standard document format, the group has chosen ODA, and its corresponding interchange format, ODIF. Among other things, this would allow a user to send a formatted text via the X.400 e-mail standard to a remote METAL system.

User-driven

The translation department of Mercedes-Benz has been working in close collaboration with the fellow TWB participants for the entire trajectory of the project. As Mercedes-Benz' Monika Hoege put it, TWB adopted a user-driven approach, involving the user as equal partner in the software lifecycle. As a test-bed for the software, her group could perceive that TWB was on the right track, yet was also able to uncover both functional and "source-code problems" (a new euphemism for bugs?) which then could be communicated to the development teams.

As part of the project, members of the AI Group at the University of Surrey worked with Fraunhofer to create a termbank for Mercedes-Benz of over approximately four thousand entries per language in English, German, and Spanish in the field of automotive engineering. The Surrey group used MATE (Machine Aided Terminology Elicitation), its suite of integrated tools for developing and managing large multilingual termbanks, to elicit the termbank from a multilingual text corpus of over 800,000 words. "The TWB project has demonstrated the importance of a corpus-informed approach to the development of term banks," summed up group leader Khurshid Ahmad. "Have lots of contexts," he said, citing corpus guru John Sinclair, "and the best definition will come up on its own."

TWB part two

Now, at the end of the first phase of the project, prototypes of the workbench running under UNIX and DOS are complete, and products based on these prototypes are being readied for market. The TWB project was recently approved for continuation under phase three of the ESPRIT program; the next two years will see further product development and market studies. The huge translation service of the CEC in Luxembourg will also be involved, to ensure, as Triumph Adler's Marianne Kugler puts it, "contact with future users who will hopefully be the potential market for the 'Translator's Workbench' as a product." Further work will continue on individual modules, such as a Macintosh version of the Greek spellchecker, and the system will undergo further beta-testing. Much work will inevitably remain in refining and polishing the software, particularly in the case of the proofing modules (where the user-interface is a crucial element), to make twb a truly attractive alternative for professional translators.

The TWB group will face substantial challenges in bringing its respective products to market; it is here where American companies tend to excel with their brazen marketing techniques. Given the murderous competition in the PC market, Olivetti, parent company of Triumph-Adler, may regard it as a way to help sell its PCs through hardware/software bundles.

Surveying the project as a whole, Toon Lowette commented, "I'm extremely impressed with TWB. It seems to be a very healthy collaboration. The partners did not set their sights on the dubious goal of MT, they tried to solve pragmatic issues such as terminology management and lookup, document comparison, and format preservation in a unified way."

TWB Participants

TA Triumph-Adler ag, Fraunhofer Society, iao, L-Cube, Mercedes-Benz ag, Siemens-Nixdorf ag, Siemens-Nixdorf cds, University of Barcelona, University of Heidelberg, University of Stuttgart, University of Surrey