

# MT Summit Looks Back to the Future

The machine-translation community celebrated 50 years of "Machine Translation: Past, Present, and Future" at the recent MT Summit held in San Diego, bringing together 314 enthusiasts from the early pioneers to tomorrow's vendors and developers. We asked Sharon Denness, probably the youngest party-goer at this year's historic MT Summit, to share her raw output on the event.

## A New Era—But Where's the Funding?



Sharon Denness

driven by the twin impact of Internet technology and the collapse of the former Soviet Union. The role of the Internet inevitably became a common theme throughout the Summit and a growing number of MT developers have already developed Web-based applications. But the enormity of the Internet and the subject domains covered could prove a real terminology problem for effective machine translation. Petke ended on a more realistic note: current funding, he reminded everyone, was still precarious.

The conference started on an enthusiastic note, with Ted Petke from the National Air Intelligence Center giving the government line. His organization has widely applied machine translation in their office environment and they clearly find it advantageous. The American government will apparently continue to develop, publicize, and use machine-translation applications. Petke spoke of a "new era for MT,"



## What's Shipping

The past is interesting but the present means business. A record 26 exhibitors plied their wares at this year's conference, with the most original novelties being systems for Chinese and Arabic. *Citac* is a Chinese-to-English PC-based translation system which has taken 16 years to develop. It features batch and interactive mode and claims to translate 1,000 words per minute. Markets for the system include China, Taiwan and Singapore. *Citac* is now developing an English-to-Chinese version.

*Globalink* is currently on a high. Their MT product is hitting the Brazilian market this year and sales have surpassed all expectations, with 5,000 copies sold at a single show. In May 1997, *Globalink* was the top-selling retail MT software package, according to the company.

Although it has always been criticized for the quality of its output and is often regarded as a light-weight in the MT stakes, *Globalink* does seem to be bringing MT to the masses, often via the Web. Perhaps it has finally shaken off its legacy as a lightweight among MT aficionados. On second thought, perhaps not: according to a member of the development team, "Globalink is a system where you can add the coverage." Perhaps they mean "value."

For localizers, software firm Corel has developed a system called *Catalyst*, which provides a single environment for producing localized versions of software. The usual approach to product localization is somewhat fragmented, with translators and software engineers using different tools to localize and

recompile the software package. *Catalyst* bundles a number of features which will make the process of localization easier and faster. These include project-management features that allow the user to track localization projects: It is possible to examine project executable files in great detail and quantify the number of project files and objects, e.g., string tables, dialog boxes, menus, and other resources. The Leverage Expert allows you to extract translated text from previous localized versions of a software product, i.e., translation memory. The Pseudo Translate Expert can be used to assess whether the translation changes spacing considerations.

## MT on the War Path

The military has always been a great believer in MT, especially the Pentagon. The Forward Area Language Converter (FALCon) was developed by the American defense section and is currently used by US Forces in Bosnia to assess the military significance of documents and to determine whether they should be translated. Six prototype systems are currently used in Bosnia by the Army's V Corps Forces and Special Operations Forces to translate documents from Serbian and Croatian into English. There

is also a Russian version of the system. The core machine-translation engine is an adapted version of Systran, and includes a laptop, paper scanner, and multiple power sources.

Plans include expanding the languages available to Arabic by summer of 1998.

### CompuServe on All Channels—But Which System?

CompuServe has had such a successful year in their MT activity that they are expanding the number of forums which will be translated. Mary Flanagan, CompuServe's MT guru, predicted that more translation of Spanish and Portuguese would be required, which suggests that Latin America will be the future market for this technology—as it is for everything else.

More excitingly, CompuServe is thinking about changing its MT system. It currently uses Transparent Language Inc.'s *Transcend* which has not proved flexible enough to add new language pairs, even though its translation speed remains unmatched. So the big question is: which system will CompuServe choose?

In another move, CompuServe is moving into spoken-language translation, with a system designed to be used in online chat forums. Users would talk into their computer and have it translate online into the required language and display the results on-screen. Using current dictation software, speakers typically pause between words—rather inappropriate for a chat forum. CompuServe plans to test the system with continuous-speech technology—a “chat” implementation may not be that remote.

In addition to the CompuServe spoken-language project, Carnegie Mellon University was showcasing *Diplomat*, described as a “rapid-deployment” speech-translation system—referring to the development of a system that performs initial translations at a useful level of quality. The CMU have developed Serbo-Croatian/English and Spanish/English bidirectional MT systems and are at work on Haitian-Creole/English and Korean/English versions.

### DEMOCRATIZING MT COVERAGE

CMU's new language pairs point to a key issue that arose during the Summit: how many languages will ever be covered by machine-translation systems? Relatively few languages are available on MT systems, with most work concentrated on the “big five”: French, German, Italian, Spanish, and Japanese, each usually paired with English. Both Harold Somers (from Manchester) and Bente Maegaard (from the Center for



Language Technology in Denmark) discussed the need for reading less widely spoken languages for MT systems. Even though the first concern in Europe is about such languages “of lesser diffusion” as the Scandinavian family, or those in Central and Eastern Europe, the debate ultimately leads to the problem of how the languages of developing nations can be factored into the MT equation, when the majority of these populations have no access to the relevant technology.

### NAMES WORTH NOTING

Lant and Alis Technologies are worth monitoring closely. They both provide *embedded* rather than *system* solutions, and there was a consensus among many speakers that MT will increasingly be integrated into other office-automation products.

Lant has been in existence a scant two years, but Managing Director Herman Caeyers has extensive experience in the field, notably on METAL, the MT system originally funded by Siemens and now recently bought by Lernout & Hauspie. Lant works in multilingual documentation and communication, using its project-oriented approach as a service-marketing point. The firm solves customers' problems by integrating and tailoring Lant's systems to customer needs. Lant's suite of tools includes LANT\*MASTER, a controlled-language checker which is based on the SECC tool developed at the University of Leuven (Belgium); EuroLang Optimizer, the translation-memory system originally developed by French company SITE, and acquired in April 1996; and LANT\*MARK, formerly known as

METAL (acquired in 1996). The languages covered are English into German, French and Spanish; French into English and Dutch; German into English, French and Spanish; Dutch into French. Lant will be launching their Lantscape system which is an Intranet infrastructure, enabling end-users to obtain translations of a number of different documents.

Alis Technologies has been developing an Arabic MT system. Its integrated MT solutions target global corporate communications, focusing on such activities as Web and intranet publishing, Web browsing, and corporate documentation. Clients include Xerox, NeocorTech, Toshiba, Transparent Language, Triad Data, Microstar, and Systran.

---

*Sharon Denness is a Research Analyst working with Equipe Consortium in the UK. You can reach her at [sharon.denness@equipe.co.uk](mailto:sharon.denness@equipe.co.uk).*