

## CoRTeXT: MT SYSTEM WITH A HEART OF GOLD

Vermont-based Universal Micro Systems Inc. announces CoRTeXT, an Intel 486 UNIX-based machine translation system. According to developer and company co-owner Stephan Bolt, CoRTeXT is capable of churning out literary translations from English to Portuguese at the amazing speed of one page per second.

Even more astonishing, perhaps, is Bolt's claim that the system is not primarily intended for commercial marketing but to "benefit mankind."

"My partner and myself are both practicing Yogis," says Bolt. "We're not in the market to make a killing. We chose to develop translation systems for the Portuguese and Indic languages not because of their commercial potential, but because they will meet real needs among the speakers of these languages."

At the moment, only the English-Portuguese module is fully operational. Other languages are under development, including French, Spanish, German, Russian, Chinese, Japanese, Indonesian, Bengali, Hindi, and Sanskrit.

CoRTeXT is starting life as a one-way

translator, but its developers are already working on an open source language module that will allow two-way translations. "We use the direct transfer translation method," says Bolt. "But we intend to use Sanskrit as an intermediate language for the Indo-European languages."

"Our approach is concept-oriented," he explains, "and Sanskrit obviously offers the greatest source of useable cognate forms." If all goes according to plan, the future CoRTeXT system will translate every one of its source languages into Sanskrit, and then out of Sanskrit and into the target language.

CoRTeXT resides on a mid-range parallel processor computer, requiring 48 Mb of RAM. Text to be translated requires extensive pre-editing, including spell-checking, and grammatical correction, and the removal of formatting codes. The "cleaned-up" text is then passed through the system for the first time.

The pre-editing software produces a concordance wordlist of terms in the text with matching translation terms from a dictionary of 14,000 terms. It also lists

words that are new to the system and require a matching term in the target language. This list is then submitted for review to the agency which has commissioned the translation.

Once terms have been checked, the text can be translated. This stage is fully automated and very rapid, reaching speeds of 3,600 pages per hour, according to Bolt. This is followed by a final post-editing phase. CoRTeXT's developers have targeted a minimum goal of 80% completion before this final stage. For technical documentation, they are hopeful that it could reach 100% completion with no post-editing required.

Bolt, a distinguished computational linguist who has worked in no less than 46 languages, stresses that the project is still in its infancy. "I hope it will be mostly used for literature, poetry, and other life-enriching texts," he says.

*Universal Micro Systems Inc., Route 100, Waitsfield, VT 05673, USA. Tel: +1 (802) 496 5500*