

[*New Scientist*, 22 February 2005]

## **Software learns to translate by reading up**

18:26 22 February 2005

Will Knight

Translation software that develops an understanding of languages by scanning through thousands of previously translated documents has been released by US researchers.

Most existing translation software uses hand-coded rules for transposing words and phrases. But the new software, developed by Kevin Knight and Daniel Marcu at the Information Sciences Institute, part of the University of Southern California, US, takes a statistical approach, building probabilistic rules about words, phrases and syntactic structures.

The pair founded a company called Language Weaver in Los Angeles, US, to sell the software as an automated translation tool. They already offer technology that can translate to or from English with four languages - Arabic, Chinese, French and Spanish.

The key to their "statistical machine translation software" are the translation dictionaries, patterns and rules - translation parameters - that the program develops. It does this by creating many different possible translation parameters based on previously translated documents and then ranking them probabilistically.

### **Genuine insights**

Knight told the American Association for the Advancement of Science meeting in Washington, DC, US, that this represents a prominent new trend in machine learning. He says the approach will eventually give computers the ability to produce genuine insights into the structure of different languages.

"What's going on now that is really exciting is translation based on syntactic structure," Knight told the meeting. "Before long a machine will discover something about linguistics that only a machine could, by crunching through billions of words."

The translated documents used to teach the translation algorithms can be electronic, on paper, or even audio files. Knight says the system is not only faster than other methods, but also better suited to tackling less common languages and the unusual vocabulary found in specialised or technical texts.

"It's an amazing process," says Benson Margulies of BASIS Technologies, a translation consultant firm based in Cambridge, Massachusetts, US. "It's the up-and-coming technology."

"The secret to machine translation is computer power," Knight adds. He says this is the only obstacle to developing even more powerful and effective translation technology. "It takes really big and fast computers."