IMPACT PROFILE
RESEARCH AT THE UNIVERSITY OF MARYLAND

AMY WEINBERG

Machine Translation of Language

Consider the following sentence from a Chinese newspaper: 穆夏拉夫在一九九九年十月的政变中取得政权, 随后一直拒绝下台。 A commercial translation system developed over many years interprets this as, "The solemn summer 拉夫 obtains the political power in October, in 1999 coup d'etat, afterwards continuously refuses to leave office."

One imagines the words are there, but the translation is puzzling. A machine translator developed at the University of Maryland in a matter of months gives a better translation: "Musharraf came to power in the coup in October 1999, has refused to step down." This is intelligible, but still not as fluent as a human's interpretation: "Musharraf obtained power in a coup in October 1999 and has refused to step down since then."

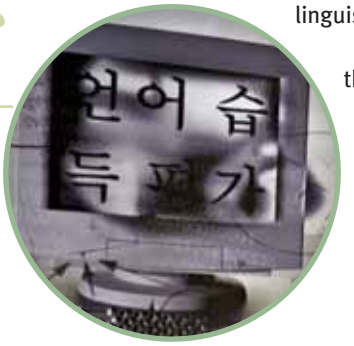
AMY WEINBERG

Language and Technology

A competent human is clearly better than a machine in translating languages, but the machines are improving, thanks to research at the University of Maryland that combines the skills of computer scientists with the expertise of linguists. "I couldn't do what I do without this interdisciplinary environment," says Amy Weinberg, an associate professor of linguistics with a joint appointment in the University of Maryland Institute for Advanced Computer Studies, or UMIACS. Weinberg also holds an appointment in the university's Center for Advanced Study of Language.

Since the 1980s, people have taken two main approaches to machine translation. "When I first got out of graduate school, putting together a machine translation system involved an army of linguists who would come up with the rules for a given language," says Weinberg. Computers would then apply the rules that linguists painstakingly derived about sentence structure and word meaning. "That approach is only as good as the number of rules a linguist can put in," she says. Commercially developed systems were often "rule-based" like this, and making them into useful systems typically consumed years of effort, whether by linguists or software engineers.

Weinberg and her colleagues have instead focused on the idea of "statistical machine translation," where rather than feeding a machine handwritten rules, the machines absorb large amounts of text that exist in two languages—such as Chinese newspaper stories and their English translations—and come up with their own rules using machine learning techniques. This idea opens up many more languages to translation—and also results in more robust systems—because the machines are trained on large quantities of real text, rather than being limited to the cases a linguist can think of.



Clearly, translation requires more than matching one word to another. For example, in Chinese, verb tense isn't clear until one reaches the end of a sentence. The key, Weinberg and her coworkers realized, is teaching computers to understand syntax.

In the late 1990s, University of Maryland researchers, with funding from the Department of Defense, started programming machines that could incorporate sentence context so that the translation would include the correct sense of each word as it is used in a given sentence. In particular, the systems learn to identify phrases—the subparts of a sentence that seem to have integrity and move as a unit. "That is a major challenge—can we define phrases?" asks Weinberg. "That challenge stopped others from taking a more complex approach."

The research at Maryland has benefited from the close collaboration of linguists with engineers and computer scientists. In contests to translate languages such as Chinese and Arabic, sponsored by the National Institute of Standards and Technology, the University of Maryland translation machines have performed very well, and their Spanish system performed as well as handwritten systems built with much greater effort.

Machine translation will probably never entirely supplant human translators. In fact, Weinberg says the question is now becoming, "How can we help humans do their work better and faster?" Machine translation could be used to scan newspaper reports and to help people search for information in English and to retrieve the answer from a translated foreign language document.

Weinberg and David Doermann, an associate research scientist in the Language and Media Processing Laboratory, are also working to make dictionaries electronically searchable. These computerized ways of absorbing and translating language can open access even to languages that are rarely used in the United States. If a political crisis suddenly makes a new dialect or language crucial to understand, newly designed machines could be invaluable assistants.

—Karin Jegalian

UNIVERSITY OF MARYLAND

Impact Profile is a supplement to *Impact*, a quarterly research digest from the University of Maryland. To learn more about research at Maryland, go to www.umresearch.umd.edu.

Produced by the Office of University Publications for the Vice President for Research, University of Maryland.