

DARPA's MT Thrust

Dragon Systems is one of five participants in a new three-year, DARPA-fueled Machine Translation project which commenced last summer. The company currently working on a prototype for a Japanese-English system called LINGSTAT. As the name suggests, LINGSTAT will be in large measure based on the corpus-based statistical techniques which Dragon engineers have developed for the language models implemented in their speech recognition system. Like DragonDictate 30k, LINGSTAT might also incorporate some form of user disambiguation, Baker suggests.

IBM is another participant in the program, with researchers at its Hawthorne, NY, lab developing a French-English MT system, called Candide, also based partly on statistical techniques. The third group involved is a consortium of researchers from New Mexico State University, the Information Science Institute of the University of Southern California, and Carnegie-Mellon University, who are collaborating on a Spanish-English system, called Pangloss, along more "traditional" AI lines.

Charles Wayne is DARPA's Program Manager for this and other spoken and written language projects in which DARPA is involved. "These are three separate projects," he emphasizes. "The only things which they have in common are that they all have English as a target language and that we will try to evaluate them in a consistent way. The first annual evaluation will be at the end of the year, but it will be something of a practice run for us."

Explaining the stimulus for the project, Wayne says it is a part of DARPA's Strategic Computing Program (although one might wonder what is strategic these days). "There was a sense that the technology had matured and was worth another look. We are trying to stimulate some innovative new approaches." Perhaps the underlying motives for the Program were best expressed in the call for proposals for DARPA's TIPSTER text retrieval and understanding project in 1990: "Proposed research should investigate innovative approaches that lead to or enable revolutionary advances in the state of the art, not focus on a specific system or hardware solution."