

## **Establishing Performance Baselines for Text Understanding Systems**

Beth Sundheim  
Naval Ocean Systems Center

The Naval Ocean Systems Center (NOSC) has been supporting DARPA's natural language research program in the area of evaluation. The objective of the work is to devise an evaluation plan that will satisfy the needs for (1) characterizing the current state of the art in theory- and implementation-independent ways and (2) setting baselines against which progress can be measured.

A task-oriented evaluation of text understanding systems was prepared and conducted. Nine different NLP systems participated in the evaluation. NOSC collected 150 texts to be used as development (i.e. training) and test data and prepared explanatory documentation on them. The performance task--a simulated database update task--and the expected outputs for each text were defined. A scoring system was devised and underwent considerable revision in the course of the evaluation.

The evaluation was conducted in phases between March and June, 1989, concluding at NOSC with the Second Message Understanding Conference (MUCK-II), at which each system was required to participate in a small onsite test. The results of the evaluation were encouraging. Some systems were able to fill a high proportion of the simulated database with very high accuracy, indicating an ability to adapt to a new domain in a short time and carry out at least a limited real-life task. A test report on the evaluation is being prepared, and tentative plans for another evaluation are being made.