INFORMATION EXTRACTION SYSTEM EVALUATION

Beth M. Sundheim

Naval Command, Control and Ocean Surveillance Center
RDT&E Division (NRaD), Code 444
San Diego, CA

PROJECT GOALS

This year, project efforts are focused on reapplying and revising existing evaluation techniques for the purpose of evaluating English and Japanese information extraction systems in the joint ventures and microelectronics domains. This year's effort will culminate in the Fifth Message Understanding Conference (MUC-5) in August, 1993.

RECENT RESULTS

MUC-4: The MUC-4 evaluation was conducted in FY92, the conference was held in June, 1992, and a proceedings was published in September. A single-value metric based on recall and precision was developed, and statistical significance tests were conducted. A blind test of 17 seventeen systems was conducted using an improved version of the Latin American terrorism information extraction task originally defined for MUC-3. Higher levels of performance by nearly all veteran systems were achieved for MUC-4, but the top scores are still only moderate. Progress in controlling the tendency to generate spurious data was obvious, but the problem still exists, along with the problem of insufficient domain coverage and general world knowledge. The push to extend the systems has brought into the focus the adverse effect that errors made in early stages of processing at the sentence and phrasal level have on suprasentential processing done in subsequent stages.

TIPSTER INTERIM EVALUATIONS: The scoring software used for MUC-4 was rewritten for the object-oriented Tipster template design. Accomodations were made for scoring Japanese. Alternative scoring procedures and new metrics were introduced. The Tipster English and Japanese systems were evaluated in September, 1992 on joint ventures, and they were evaluated in February, 1993 on both joint ventures and microelectronics. The results of these evaluations are being used to make decisions concerning the evaluation methodology to be used for the final Tipster evaluation (which will be the MUC-5 evaluation).

MUC-5: The call for participation in MUC-5 was issued in October, 1992, and participants began

development in March, 1993, in preparation for the evaluation in July and the conference in August. Over 20 organizations (including Tipster-sponsored organizations) are planning to participate. Most of the non-Tipster organizations will be working only on the English joint ventures task or the English microelectronics task; however, two will be working on joint ventures in both languages, and one will be working on microelectronics in Japanese only.

PLANS FOR THE YEAR

• Improve the evaluation methodology to be used for MUC-5 based on the experiences of the Tipster interim evaluations.

• Coordinate the MUC-5 evaluation and conduct the conference.

• Foster interest in resource-sharing among evaluation participants to support future R&D on information extraction and NLP in general.