④

INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE
DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT
Unisys/Defense Systems

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.
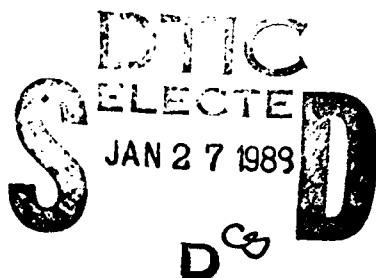
1. Description of Progress

Progress has been severely hampered by the funding situation. Our current funding authorization ran out at the end of June. Since then we have been authorised by Unisys to proceed at risk to a limited extent, which has supported only a maintenance level of effort on the contract.

1.1. Evaluation of NL systems

Martha Palmer has organised a workshop on the evaluation of natural language systems sponsored by AAAI and RADC, which will be held on December 8 and 9. The program committee met in Paoli to select abstracts for presentation at the workshop and define the schedule. 18 abstracts were accepted. A number of additional participants were invited, bringing the total number of attendees to 46.

1.2. Evaluation of PUNDIT

We performed an intial black box evaluation of Pundit using a simple message processing/information retrieval task. RAINFORM sighting messages were first processed by Pundit and stored in a relational database format. The database was then queried and messages which answer the query were retrieved. It is relatively easy to define correct performance for this task in terms of recall and precision measures in comparison with human information retrieval performance. For example, if a human understands a message to refer to the sighting of a submarine by a ship, then a query such as "Did any ship sight a submarine?" should retrieve that message (this contributes to the score for recall), and not retrieve messages that do not describe sightings (this contributes to the score for precision). A set of 17 RAINFORM sighting messages was used as a training set and 17 additional messages were used as a test set. Human responses were collected from two responders for both sets of messages. Initial measurements indicate that average Pundit performance for the training set is about 16% in recall, and 100% in precision. On the test data, Pundit's recall was about 6% and precision was 100%. In essence, Pundit retrieves very few of the relevant messages, but everything that it retrieves is relevant. An analysis of the reasons for failure was performed. We plan to fix some of these problems and run the evaluation again.

1.3. Resource Management Domain

A merged lexicon for all domains has been created. The merged lexicon contains about 2400 distinct root words. A tool for extracting lexical items from a dictionary, given a word list,was used to obtain items from the merged lexicon for the resource management domain. Approximately 47% of the vocabulary (or 448 words) in this domain is new. However a large number of these new words are proper names of ships and locations, which are relatively trivial to enter. There are about 225 other words, including about 50 verbs. Because there are so few verbs, we expect adding the lexical entries to take only a few days.

2. Change in Key Personnel

3. Summary of Substantive Information from Meetings and Conferences

3.1. Darpa Meetings

Lynette Hirschman and Deborah Dahl attended the meeting on performance tasks for spoken language systems, September 7-8, in Washington.

Charles Wayne met with the natural language group in Paoli on October 21. We demonstrated the various components of Pundit and discussed evaluation, portability tools, and our spoken language efforts. The results of the black box evaluation of Pundit were presented at this meeting.

3.2. Papers and Presentations

Martha Palmer has completed revisions to her book, which is in press at Cambridge University Press, on the semantic interpreter used in Pundit. This will provide the basis of documentation for the semantics portion of the system.

(1) Ball, C. Pragmatic Change in the Development of the NP-Focus Cleft. Accepted for presentation at the Annual Meeting of the Linguistic Society of America.

(2) Dahl, D., Hirschman, L., and Ball, C. A Black Box Evaluation of the Pundit System. To be presented at the Workshop on Evaluation of Natural Language Systems, Dec. 7-8, 1988.

(3) Hirschman, L., Hopkins, W., and Smith, R. "Or-Parallel Speed-Up in Natural Language Processing: A Case Study," Proc. of the 1988 International Joint Conference on Logic Programming, Seattle, WA, August 15-19, 1988.

(4) L. Hirschman, "A Meta-Treatment of Wh-Constructions in English", Proc. of META'88, Meta-Programming in Logic Programming, Bristol, UK, June, 1988; also to appear in book form, edited by M.A.H. Rogers and H. Abramson, 1989.

(5) Hirschman, L., Palmer, M., Dowding, J., Dahl, D., Linebarger, M., Passonneau, R., Lang, F., Ball, C., and Weir, C. The Pundit Natural Language Processing System. Submitted to AISG (Artificial Intelligence in Government).

(6) Passonneau, R. A Computational Model of the Semantics of Tense and Aspect. *Computational Linguistics,* 14-2, June, 1988, pp. 44-60.

3.3. Conference Attendance

Lynette Hirschman attended the 5th International Logic Programming Conference, Seattle, August, 1988.

4. Problems Expected or Anticipated

Authorized funding ran out in June. The additional increment for FY 88 of $86K and the FY 89 funding of $412K are about to be released out of ONR. This will permit us to resume work at our normal level of effort.

5. Action Required by the Government

Sign off on final $14K increment.

6. Fiscal Status

(1) Amount currently provided on contract:
$ 1,192,833 (funded) $1,704,901 (contract value)

(2) Expenditures and commitments to date:
$ 1,223,040

(3) Funds required to complete work:
$ 481,861