

THE CONSORTIUM FOR LEXICAL RESEARCH

Y. Wilks, Principal Investigator

Computing Research Laboratory
New Mexico State University
Las Cruces, New Mexico 88003

PROJECT GOALS

Work in computational linguistics has reached the point where the performance of many natural language processing systems is limited by a "lexical bottleneck". That is, such systems could handle much more text and produce much more impressive application results were it not for the fact that their lexicons are too small.

The Association for Computational Linguistics has established the Consortium for Lexical Research (CLR), with DARPA funding at the Computing Research Laboratory. The Consortium is advised by an ACL steering committee consisting of Roy Byrd, Ralph Grishman, Mark Liberman and Don Walker.

The task of the CLR is to facilitate research, making available to the whole natural language processing community certain resources now held only by a few groups that have special relationships with companies or dictionary publishers. The CLR intends that agreements are set up with "providers" on reassuring and advantageous terms to both suppliers and researchers. The CLR will as far as is practically possible accept contributions from any source, regardless of theoretical orientation, and make them available as widely as possible for research.

RECENT RESULTS

The Consortium project officially started in January 1991. Since then, an infrastructure for distribution and receipt of materials has been set up and contacts with researchers and major publishers have been established. We have now reached a stage where memberships are being accepted and resources are being acquired and distributed.

Our current status can be summarized as follows:

Distribution Secure ftp procedures have been set up for three kinds of materials: free, members only, and restricted (by a fee or additional contract or both)

Collection A group of public domain resources have been obtained and cataloged and acquisitions of software

and data from publishers and researchers is ongoing.

Contracts Together with our university lawyers, we have developed contracts for members and providers. These are now undergoing final revision incorporating changes suggested by several publishers.

Membership We have had inquiries from over 500 sites. The first membership agreements have now been received.

Documentation We have prepared several types of information related to the CLR holdings: A catalog of short (paragraph) descriptions of each holding, a catalog of where to find the holding (via ftp), and an individual detailed descriptions of the software and data for each holding.

Publicity The Consortium has been announced through email network news, in the ACL journal, and at various conferences in the U.S., Europe and Japan.

In January 1992, the Consortium hosted a workshop where publishers, researchers from academia, government and industry, and lexicographers gathered to discuss their needs and plans. The feedback from the workshop was positive and is now shaping the development of the CLR in its second year of existence.

PLANS FOR THE COMING YEAR

Provisional agreements have been established with Longman and Harper-Collins for allowing us to act as a broker for their electronic materials. These agreements will be finalized in the first part of this year. This will be a major milestone for the Consortium. The form of the agreements that they will sign with the members still remains to be settled. Once this is finalized, we hope this will provide a model which will allow us to complete agreements more easily with other publishers.

We plan to expand membership and holdings steadily over the year, and progress toward our long term goal of establishing the Consortium as a self-supporting entity. For information on the CLR, contact lexical@nmsu.edu.