

# THE CONSORTIUM FOR LEXICAL RESEARCH

*Y. Wilks, Principal Investigator*

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

## PROJECT GOALS

The Consortium for Lexical Research (CLR), established by the Association for Computational Linguistics, with funding from DARPA, is now beginning its third year. The Consortium is sited at the Computing Research Laboratory, New Mexico, USA, under its Director, Yorick Wilks, Associate Director Louise Guthrie, and an ACL advisory committee consisting of Roy Byrd, Ralph Grishman, Mark Liberman and Don Walker.

The objective of the Consortium for Lexical Research is to act as a clearinghouse, in the US and internationally, for lexical data and software. It shares lexical data and tools used to perform research on natural language dictionaries and lexicons, as well as communicating the results of that research, thus accelerating the scale and speed of the development of natural language understanding programs via standard lexicons and software.

The task of the CLR is primarily 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, as far as is practical, accepts contributions from any source, regardless of theoretical orientation, and makes them available as widely as possible for research.

## RECENT RESULTS

Our focus this year has been on the acquisition of new materials and the recruiting of new members. Response to the Consortium has been enthusiastic and continuous. The repository has grown significantly and the consortium membership has quadrupled in the last year. Information about the CLR including the catalog of offerings, the membership or provider agreements or any previous newsletter can be obtained from [lexical@nmsu.edu](mailto:lexical@nmsu.edu).

Our current status can be summarized as follows:

**Collection A** group of public domain resources has been obtained and cataloged and acquisitions of software and data from publishers and researchers is ongoing. We now offer nearly 100 items.

**Contracts** Together with our university lawyers, we have developed contracts for members and providers. Negotiations with dictionary publishers have been difficult, but we now have arrangements with Longman and Harper-Collins publishers which facilitate the purchase of their machine readable dictionaries by members. At present the distribution of the dictionaries is still in the hands of the publishers and is slow. We are working on ways to expedite the process.

**Membership** We now have 54 members of the CLR: 28 universities, 20 companies (including Apple, Microsoft and Xerox) and 4 government organizations. We have had over 2,000 ftp accesses in the last six months from more than 20 countries around the world.

**Publicity** The Consortium has begun a newsletter which is distributed to members and to anyone who has requested information about the CLR. The newsletter highlights a different piece of software or data each month, and informs its readers of any new items which are available.

**Conferences** In December 1992 the Computing Research Laboratory hosted the second workshop of the Consortium for Lexical Research: U.S./European Cooperation. The workshop was sponsored jointly by NSF and the European Commission to discuss international cooperation of lexical computation. Twenty-five researchers participated in the workshop. A report is available through the Computing Research Laboratory.

## PLANS FOR THE COMING YEAR

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. We hope to do this by signing agreements with other dictionary publishers to make their products available through the CLR and by more actively seeking contributions of software or data from researchers. Our membership drive will focus on obtaining more international members, and members from the community of researchers and language specialists who may not have everyday access to the internet.