

THE CONSORTIUM FOR LEXICAL RESEARCH

Louise Guthrie, 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 ending its third year. The Consortium is sited at the Computing Research Laboratory, New Mexico, USA, under its Director, Louise Guthrie, and an ACL advisory committee consisting of Roy Byrd, Ralph Grishman, and Mark Liberman.

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, 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.

RECENT RESULTS

With a repository of over 275 items, ftp accesses averaging almost 8,000 per month, and a membership roster which includes most of the major natural language processing centers in the U.S., CLR has fulfilled many of the goals envisioned at its inception. 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. 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.

Our current status can be summarized as follows:

Collection The CLR repository has grown significantly in the last year, becoming especially strong in tools for linguistic analysis. Our holdings encompass lexicons, parsers, morphological analyzers, dictionary tools, and a wide variety of other materials. We provide a descriptive catalog of all major holdings. A group of public

domain resources has also been obtained and cataloged. The acquisition of software and data from publishers and researchers is ongoing.

Contracts Together with our university lawyers, we have developed contracts for members and providers. Negotiations with dictionary publishers have been slow, but we now have arrangements with Longman and Harper-Collins publishers to facilitate the purchase of their machine readable dictionaries by members. We are currently working on contracts with two more dictionary publishers, one American and one foreign.

Membership We now have 57 members of the CLR: 30 universities, 19 companies (including Apple, BBN, and Xerox) and 8 government organizations. Membership growth is led by acquisitions development; as our repository becomes more substantial we have witnessed a concurrent rise in membership inquiries.

Publicity The Consortium publishes a newsletter 8 times a year, which is distributed to its members and to core researchers in lexicography. The current mailing list has more than 700 individuals and organizations, many of whom wrote and requested that their names be included. The newsletter highlights topical areas, and informs members of new materials.

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 continuing to solicit contributions of software or data from researchers. CLR has enjoyed a supportive working relationship with the Linguistic Data Consortium (LDC) since our inception; we hope to foster still closer ties which will include joint planning and operational activities. We are also seeking to expand CLR through alliances in Europe. This would allow quicker access to our archives for international members and open up important avenues for CLR into the acquisition of European resources.