

Title of proposed student learning community: Cognitive Science Group

Dear Professor Tohline:

We are writing to propose that the information technology initiative at LSU include creation of a computational linguistics group within the planned LSU Center for Applied Information Technology and Learning. The research focus of this group would be the automatic processing of natural language, including automatic perception, parsing, interpretation, and production of utterances of English and other languages. We believe that such a research and educational group within the Center would

- help create a workforce in Louisiana with relevant skills for the information age economy;
- together with existing programs, put LSU at the forefront of international efforts in information technology related research;
- foster the creation of public-private partnerships between LSU faculty and private industry;
- help faculty retention, by helping existing faculty computationally model their theories, and by creating a more exciting research environment; and
- attract and retain high quality students from in and out of state.

The task of facilitating human-computer interaction is widely considered to be the most important factor in making digital information technology more widely usable within society. Errors, delays, and repetitive stress injury are among the consequences of the mismatch between human and computer capabilities. Computational linguistics addresses these difficulties by building systems that can automatically process and interpret human spoken and written language, so that human-computer interaction can proceed on grounds on which people are most comfortable and effective.

As business and research become progressively more internationalized, there is a need to make information available in a wide variety of languages. This has led to the development of international character sets for word processing and for the world wide web, and to research into automatic translation systems. Computational linguistics is at the forefront of these efforts.

Finally, computational linguistics is centrally involved in advances in information storage and retrieval. Its contributions to this area include annotating and tagging large corpora for various aspects of linguistic structure in order to facilitate the automatic extraction of information.

In considering the central role of computational linguistics in contemporary technology, one need only consult the latest issue (as of September 23) of the LINGUIST list, (<http://www.emich.edu/~linguist/jobsindex.html>) which advertises over 100 jobs in computational linguistics. Companies and organizations such as Microsoft, Xerox, Infotech, SRA International, TCC, BCL Computers, the Educational Testing Service, and Cognitive and Communication Technologies advertise for project management and programming positions involving speech recognition, world wide web development, human language interfaces, natural language processing, and knowledge extraction. Listings for computational linguists formed the largest single area of expertise advertised. As industry reports testify, (e.g. "Wordsmiths of the Web," Wall Street Journal, 2000),

computational linguistics will be vital to the development of digital technology for decades to come.

If Louisiana residents interested in digital technology for education and careers are going to stay in-state, and if information technology firms are to be based in Louisiana, areas of research and technology involving computational linguistics must be well represented in the research and educational efforts at research institutions within the state. They are currently not well represented; in fact, they are hardly represented at all.

Examination of the twenty three faculty members listed on the LSU linguistics web page (<http://www.artsci.lsu.edu/ling/>), as well as the list of thirty faculty members on the recently completed LSU cognitive science group web site (<http://www.artsci.lsu.edu/cogsci/faculty/faculty.html>), shows that many LSU scholars have research interests that intersect with computational linguistics, but that there are no specialists in this field at LSU. Neither Electrical Engineering nor Computer Science has any faculty with research proficiency in computational linguistics. Furthermore, computational linguistics is not a primary research area of the Institute of Cognitive Science at the University of Louisiana at Lafayette. That program has no computational linguistics lab (out of seven labs total), no allied linguistics program (unlike LSU), and only one scholar (Sheri L. Condon) with computational linguistics expertise.

Unlike LSU and ULL, most major research universities in the US and Western Europe support work on language structure, use, acquisition, and impairments, accompanied by allied computational studies, involving computational linguistics or computational modeling of linguistic phenomena. For example, at Stanford's Center for the Study of Language and Information (CSLI), the study of language structure, acquisition, and use goes hand-in-hand with a wide variety of computational studies in the context of a major academic-industry consortium at Xerox Park. At many other institutions, such work is housed entirely within the university. For example, studies in language and language acquisition are intimately linked to work in language technologies in Carnegie Mellon's Psychology and Computer Science departments, at Penn's Institute for Research in Cognitive Science, in John Hopkins' Department of Cognitive Science, and Brown's Department of Cognitive and Linguistic Sciences. In Europe, programs of this sort are proliferating under the impetus of the European Union.

At LSU, many world-class research efforts into various aspects of language use, acquisition, and impairment are currently underway within traditional departments or in the Interdisciplinary Program in Linguistics. But computational studies that could, and at many universities would, be allied with these efforts are notably absent at LSU. There is extensive cutting edge work on language acquisition at LSU, by researchers in the Department of Psychology and the Department of Communications Sciences and Disorders, yet no work on computational modeling of language acquisition, a rich field employing a host of computational techniques, including connectionist, rule-based, and evolutionary systems. There is much work on the phonetics of human speech at LSU, but no allied work on automatic speech perception. There is extensive and high-quality work on speech disorders and deficits, but no allied work on computational modeling of such conditions. There is excellent work on linguistic semantics and the philosophy of language, but no allied work on computational semantics, a rich field employing a wide array of computational approaches, including both connectionist and non-connectionist

versions of statistical processing, as well as rule-based systems. There is excellent work at LSU on expert systems, information retrieval, and human-machine interfaces in Computer Science, Electrical Engineering, Industrial and Manufacturing Systems Engineering, and Information Systems and Decision Science, yet no specialists in either principle-based syntactic processing or corpus based/statistical processing of natural language.

We contend that the State of Louisiana will be missing out on a significant part of future work on digital information technology and processing if it does not attend to the human-computer interface by constituting a group in computational linguistics to address issues in natural language processing. Furthermore, excellent research programs across the campus are in danger of losing ground relative to their national and international peers due to the lack of allied computational pursuits on campus. We therefore suggest that the LSU Center for Applied Information Technology and Learning include an array of positions in computational linguistics, with something like the following breakdown (the exact composition of the group could be decided by one or two senior researchers hired initially):

- Computational models of language acquisition (1 position)
- Computational linguistics: principle-based syntactic processing (2 positions)
- Computational linguistics: corpus based/statistical processing (1 position)
- Computational semantics: reference, discourse representation, translation  
(3 positions)
- Automatic speech perception (1 position)

There are many possibilities for housing these appointments. We think that most of these appointments would **not** be most effective if they were housed exclusively in a traditional academic unit, but that all of them could usefully interact with researchers in traditional units. Thus, we think that these appointments would be best if they were appointments directly to the new Center for Applied Information Technology and Learning (CAPITAL), or if they were joint appointments between the CAPITAL and more traditional departments. An appointment in computational models of language acquisition could be a joint appointment between CAPITAL and Psychology or Communications Disorders; appointments in computational linguistics (principle-based syntactic and corpus-based/statistical) could be joint appointments between CAPITAL and Computer Science; appointments in computational semantics could be joint between CAPITAL and Philosophy; and appointments in automatic speech perception could be joint between CAPITAL and Psychology or Communication Disorders.

A research group such as that sketched here would not merely allow LSU to play “catch-up” to peer institutions, but rather, with existing programs, place LSU at the forefront of information technology research and applications relating to computational linguistics.

Signed,

Jon Cogburn  
Assistant Professor  
Department of Philosophy and Program in Linguistics

Michael Hegarty  
Assistant Professor  
Department of English and Program in Linguistics