Motivation and Goals

Ron Cole & Mark Liberman

The world is changing at a remarkable pace. Not too long ago, we added fax numbers to our business cards. A few years ago, we added our web address. The growth of the internet is changing the way we work, communicate and play in fundamental ways. For example, the Language Resources and Grantees Workshops were organized entirely by email and the Web.

These changes have a profound impact on our national agenda. Access to the internet by citizens to learn, communicate, create and publish has become a national priority. As the nature of communication changes, the nature of surveillance and intelligence gathering must change to protect our nation. As the cold war wanes and new threats emerge, the mission and needs of our armed forces must change as well.

Human language technologies play a vital role in each of these areas, enabling people to request, access, locate, summarize, receive, organize, understand, act upon, create and communicate information. Advances in human language technology are needed to enable these activities, and these advances require new language resources. At the same time, technological advances enable both researchers and publishers to create and distribute language resources in new and more efficient ways, while intellectual property rights come to the fore as a problem for effective use of these resources in a networked world.

What are language resources? The preliminary announcement of the First International Conference on Language Resources and Evaluation (Granada, Spain, 28-30 May 1998) has this to say:

> The term language resources (LR) refers to sets of language data and descriptions in machine readable form, used specifically for building, improving or evaluating natural language and speech algorithms or systems, and in general, as core resources for the software localization and language services industries, for language studies, electronic publishing, international transactions, subject-area specialists and end users. Examples of linguistic resources are written and spoken corpora, computational lexicons, grammars, terminology databases, basic software tools for the acquisition, preparation, collection, management, customization and use of these and other resources.

The relevance of evaluation in Language Engineering is increasingly recognized. This involves assessment of the state-of-the-art for a given

technology, measuring the progress achieved within a program, comparing different approaches to a given problem and choosing the best solution, knowing its advantages and drawbacks, assessment of the availability of technologies for a given application, and finally product benchmarking. It accompanies research and development in Human Language Technologies, and has driven important advances in the recent past in various aspects of both written and spoken language processing. Although the evaluation paradigm has been studied and used in large national and international programs, including the US ARPA HLT program, EU Language Engineering projects, the Francophone Aupelf-Uref program and others, particularly in the localization industry (LISA and LRC), it is still subject to substantial unresolved basic research problems."

The goals of the workshop are:

-
(a) to assess the needs of the research community for language resources, and identify key areas of human language technology in which new resources are essential for research advances to occur;

-
(b) to review current efforts to develop and distribute language resources, and determine how well these efforts are meeting the needs of the research community and the federal agencies that support research and development of human language technologies and systems.

-
(c) to identify and propose new initiatives and models to meet the needs of the research community and its sponsors.