KNOWLEDGE-BASED APPROACHES: SESSION INTRODUCTION

Ivan A. Sag
Department of Linguistics
Stanford University
Stanford, CA 94305

As research in natural language processing spawns commercial products, as it now begins to, our research community acquires new ethical and practical responsibilities. All software packages have limitations, and research progress consists in devising techniques that lead to reducing these limitations. There is every reason to believe that currently available natural language interfaces may be of significant value to those who purchase them, despite their limitations, and that their successors will be of even greater value. Our responsibility to accurately convey the limitations of existing systems, as well as their capabilities, is two-fold: an ethical responsibility to the public not to mislead potential consumers, and a practical responsibility to each other not to destroy the credibility of future research results that lead to enhanced language processing capabilities.

How are we living up to this responsibility? One commercially available product now advertises by depicting users that are satisfied because they can "simply type [their] question in everyday conversational dialogue, PHRASED ANY WAY [THEY] LIKE [emphasis added]". Another boasts an "understanding of UNRESTRICTED [emphasis added] English queries comparable to a human's". The former acknowledges that "every product has a limit to what it can do" (The only English sentence offered as beyond the system's capability, however, is: "To be or not to be, ..."). But the latter proudly asserts: "DISADVANTAGES: none". The former is at best misleading. But the latter makes a mockery of both the intellectual content of our discipline and the basic principles of honesty in advertising.

How do we assess the limitations of natural language systems? When we are presented with a user-computer interaction, whether in a live demonstration or in a scholarly paper such as those being presented at this conference, there is an implicit trust. We assume, for example, that each sentence being processed is really being processed, and further that each sentence is merely one instance of a large class of similar sentences, all of which can be processed in like fashion. Likewise, when we see cooperative, seemingly intelligent responses to queries, we assume that there is some generality to the intelligence, that it will exhibit itself in a variety of different circumstances, and will transport itself to different domains.

Thus we as a community would be shocked to hear reports of public demonstrations of INSTALLED natural language systems where unexpected deviations from planned presentations expose what appear to be "canned" responses rather than intelligent language processing. If such incidents were to occur, there would be great cause for alarm. The implicit trust I referred to earlier would have been violated, and the very integrity of our discipline called into question. To retain credibility as a responsible research community, we must devise better methods for evaluating research, ways of exposing hucksters and charlatans in our midst, and some means of enforcing standards that guard against overzealous and unprincipled marketing of commercial products.

Sensible research of course takes many shapes. Among the key issues of our field that make good sense to address are the following:

1. new techniques for representation or processing
2. limitations of existing techniques
3. comparisons among existing techniques
4. new problems for theories of representation or processing
5. properties of system design that ensure desirable functioning (e.g., discussions of modularity and portability)
6. assessing the limitations of existing systems

The papers in this session touch on several such issues. Boguraev and Sparck Jones discuss number 5; both Finin and Palmer and Pazzani and Engelman touch on number 3; Montgomery is concerned with number 4; Lehnert and Shwartz address number 6. In some cases the contribution being made is clear. In other cases, I for one find it difficult to assess whether the systems being described do what they do in a principled way that will generalize to other applications, or in an ad hoc fashion that is devoid of intellectual interest. In the absence of agreed-upon standards for evaluating descriptions of such systems, I can only urge both authors and audience to address the issues I have raised, doubtless not for the first time, in the oral presentations and subsequent discussion.