

THE RATE OF PROGRESS IN NATURAL LANGUAGE PROCESSING¹

Norman K. Sondheimer
USC/Information Sciences Institute
Marina del Rey, CA 90292

With all due respect, the rate of progress in natural language processing has been disappointing to many, including myself. It is not just that the popular press has had overblown expectations, but that we at this meeting have. The consequences of these errors could be severe. Hopefully, this short note will give an accurate evaluation of our rate of progress, identify what some of the problems have been, and present some reasonable suggestions on what can be done to improve the situation.

WHERE ARE WE?

The most obvious evidence of slow progress is found at the end of the chain from research through development to application. Practical natural language interfaces, writing aids, and machine translation systems all exist. But the public has not been quick to accept what we can produce. I know of no company that has "gotten rich" off natural language interfaces. More importantly, in my estimation the most technically successful natural language interface to database systems was introduced in the late 1970's. Although the research community has been quick to point out shortcomings with that system and other systems have been introduced, no clear rival has appeared. Commercial MT efforts follow the same pattern.

Moving backwards along the chain, serious large-scale prototypes of the next generation of systems are hard to find. This is not due to lack of industrial interest. All major computer manufacturers seem to have been interested in natural language processing in recent years. Those systems which I have heard about generally appear to be severely limited and habitually delayed. The next serious competitor to existing commercial products is not obvious to me.

More common are the initial laboratory demonstrations of new understanders and generators, as well as their components. Finally, at the beginning of the chain, are the ideas for new systems that come from new frameworks, new perspectives on the problem, and new insights from related disciplines. These are the stuff of our conferences and journals. Here may be found the possibility of real progress at a good pace.

Yet, even though the years since the first TINLAP have seen a steady stream of new ideas, I find no

¹This work is supported by the Defense Advanced Research Projects Agency under Contract No MDA903 81 C 0335 and by the Air Office of Scientific Research under FQ8671-84-01007. Views and conclusions contained in this report are the author's and should not be interpreted as representing the official opinion or policy of DARPA, AFOSR, the U.S. Government, or any person or agency connected with them.

I am delighted to thank my colleagues: Ralph Weischedel, Ray Perrault, Tom Galloway, Ron Ohlander, Ed Hovy, Bob Neches, Larry Miller, Bob Kasper, Mitch Marcus, Larry Birnbaum and Bill Mann, for taking the time to set me straight.

special reason to believe that these will be better able to scale up and still solve the difficult problems that have always faced us. These problems include lexical ambiguity, ill-formed input, metonymy, and even the fundamental problem presented by the size of a realistic knowledge base. Without greater proof of the ideas usefulness, they serve at best as better insights into the problems natural language presents to us. Although these may be useful to us and others who study language, they cannot be accepted as ends in themselves for a field that is defined in terms of machine processing.

If my analyses are correct, it is unreasonable to expect the broad base of support we have thus far been provided to continue.

WHAT IS WRONG HERE?

I can only guess where the problems lie and I can only do that from my personal perspective. You can assume that I have seen every one of these mistakes in my own behavior.

A fundamental problem is that I and, probably, most researchers are not truly realistic about the difficulty of the problem. Most of us do try hard to understand our situation, promise only what we think we can deliver, and do our best to develop appropriate public expectations. Even so, the problem is that we probably still underestimate the difficulties. It is likely that there is still much more to natural language than we now realize. How can we really say what we need to allow for to achieve truly human level performance? The mere fact that we take the problem to be formalizing one of the most complex human abilities may well make complete success impossible.

It is also likely that we can't hope to unambiguously identify progress. We can get neither the type of experimental evidence that physics or chemistry requires or the rigorous proofs that mathematics can produce. Given the nature of language, we must settle for carefully reasoned arguments for our proposals based on limited and challengeable insights and many explicit and implicit assumptions. In this respect, we resemble the "soft" social sciences. Fortunately, we are also like engineering in that we should be able to measure our results in terms of a body of useful techniques of limited utility characterized by appropriate case studies. That doesn't sound half bad to me; if only we were doing a good job of it!

But I think we have some serious sociological problems that keep us from making faster progress. We seem to value the most theoretically ambitious research far out of proportion to its proven worth. Such work has the best possibilities for publication and gets the most respect from our colleagues. In addition, jobs and funding aimed at achieving such results come with the least commitments. All of these are natural and good things - in limited amounts.

Consider, however, what often results. Sometimes we resemble a school of fish. When our leaders turn, many of us turn with them. Unification and connectionism are only the latest turning. We do it all the time. Heck, I do it. It's fun to work on new things; for the first few years there are lots of easy problems to solve. This schooling behavior probably happens in every field. However, it is especially bad in our case because we rarely get the old technology worked out in enough detail to really evaluate its usefulness.

A related error on our part finds us acting like "fish out of water" when we enter the worlds of the philosopher, linguist or psychologist. Naturally, we want the respect of the older disciplines that are concerned with language. However, their values can not possibly match ours very well. Unfortunately, we have often ended up adopting theirs and abandoning our own. When this happens the results of our research have less and less likelihood of contributing to the progress of our computational discipline. Concluding the fish metaphor, it is clear that in order to communicate with them, we are going to have to ask our friends in other disciplines to learn to swim with us.

I could explore some of the other problems that impede progress, such as our awful tendency to focus on solutions to particular problems without thinking through their compatibility with solutions to other problems, our studied ignorance of earlier work, our willingness to accept unproven ideas as the basis for further work, and our tradition of not warning readers of known shortcomings of our results. However, before you give up on me completely, let me suggest some future directions.

WHAT CAN WE DO?

Am I ready to give up on natural language processing? Certainly not. If I were, I would not be in my office on a perfectly gorgeous Southern California Sunday writing this. In fact, I'm more ready than ever to push on. As nice as Las Cruces and this meeting are, it's hard for me to justify being away from my work for three days. Besides, the situation is not hopeless. I'll refrain from pushing my favorite technology; instead, I'll try the trickier tactic of addressing our field's values.

Our field exists because of one natural phenomenon, human language, and one technology, the computer. Our values must come from these two roots. It is easy to see that we have to value the meanings and uses of human language in building our systems. Clearly, the ultimate goal must be to understand or generate language in a way that matches what we see humans do.

More important to point out at this conference are the values from our computational root. We have shown some concern for computational complexity, but usually of the worst case sort, not the more important average performance. But there are other concerns as well; the ease of coding an algorithm, the ease of maintaining and enhancing a system, the portability of the system, the way in which the system responds to output beyond its basic coverage, how it responds to ambiguity and vagueness, the facilities available to tailor a system to an application, site, or user, and so on. Probably, the most confusing pressure from computation comes to natural language interfaces from the fact that people end up communicating with the machine in ways that they would never communicate with other people. We must value these realities as much as we value the demands of natural human communication. Such topics should be discussed as often as anaphora, metaphor, conjunction, et al., are in our panels and papers.

Values of another sort have to come from the society that supports us. It is not just the ethics of accepting a salary; it is a matter of self-preservation. We simply have to pay more attention to pushing our own ideas down the chain from theoretical research. The outside world is not going to believe we are

making progress unless they see something come of our ideas in terms they can understand. And if the people at this conference do not see to it that this happens, who will? And if we do not do it now, when will we have the chance again?

Given that we want to take our ideas down the chain from theoretical research to empirical study and beyond AND that natural language is an extremely difficult task, how can we proceed? There is only one answer: work within our current limits. Let's treat our work as that of successive approximations. Let us forget about the unexplored problems for the time being. Let us see what we can really do with the proposals we have that seem to work. Basically, let us emphasize building systems and full-scale components for a while.

For example, why don't a group of us take the best parser, the best semantic interpreter, the best generator, the best inference system, etc., and tie them together? Then let's pick a domain of discourse and make them work for more than a few sentences. Let's beat on them until they work for as much of language as they appear capable. While we are at it, let's make the system as fast, as robust, as portable, as maintainable, etc., as we possibly can. Similarly, let's beat on individual components in the same way.

I know there is no guarantee this approach will produce a useful system or component. But even if we fail to produce something worth going further with, we will have learned a lot about what works and what doesn't. If those results are not allowed to be lost, the next effort can do better.

Of course, a problem with this approach lies in the source of our funds. Rare is the company or funding organization that is not asking for new ideas and encouraging us to move on. So we have to convince them that stability is necessary for systems building and the overall well-being of the field.

Our field arose out of a perceived need for language processing systems. The basic problem we have is that we have not been able to produce these systems at the rate we had thought possible. Unless we turn our primary attention to increasing the speed our theoretical ideas move out to initial demonstrations, initial demonstrations move out to prototype systems, and so on, we will face a serious crisis. To bring the point home, if we do not remember why the field of natural language processing exists and accept the necessary values, I venture to guess that there will be little external support for a TINLAP in the not too distant future.