REFLECTIONS ON TWENTY YEARS OF THE ACL

Jonathan Allen
Research Laboratory of Electronics
and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

I entered the field of computational linguistics in 1967 and one of my earliest recollections is of studying the Harvard Syntactic Analyzer. To this date, this parser is one of the best documented programs and the extensive discussions cover a wide range of English syntax. It is sobering to recall that this analyzer was implemented on an IBM 7090 computer using 32K words of memory with tape as its mass storage medium. A great deal of attention was focussed on means to deal with the main memory and mass storage limitations. It is also interesting to reflect back on the decision made in the Harvard Syntactic Analyzer to use a large number of parts of speech, presumably, to aid the refinement of the analysis. Unfortunately, this introduction of such a large number of parts of speech (approximately 300) led to a large number of unanticipated ambiguous parsings, rather than cutting down on the number of legitimate parsings as had been hoped for. This analyzer functioned at a time when revelations about the amount of inherent ambiguity in English (and other natural languages) was a relatively new thing and the Harvard Analyzer produced all possible parsings for a given sentence. At that time, some effort was focused on discovering a use for all these different parsings and I can recall that one such application was the parsing of the Geneva Nuclear Convention. By displaying the large number of possible interpretations of the sentence, it was in fact possible to flush out possible misinterpretations of the document and I believe that some editing was performed in order to remove these ambiguities.

In the late sixties, there was also a substantial effort to attempt parsing in terms of a transformational grammar. Stan Petrick's Doctoral Thesis dealt with this problem, using underlying logical forms very different from those described by Chomsky, and another effort at Mitre Corporation, led by Don Walker, also built a transformational parser. I think it is significant that this early effort at Mitre was one of the first examples where linguists were directly involved in computational applications.

It is interesting that in the development of syntax, from the perspective of both linguists and computational linguists, there has been a continuing need to develop formalisms that provided both insight, as well as coverage. I think these two requirements can be seen both in transformational grammar and the ATN formalism. Thus, transformational grammar provided a simple, insightful base through the use of context-free grammar and then provided for the difficulties of the syntax by adding on to this base the use of transformations and of course, gaining turing machine power in the process. Similarly, ATNs provided the simple base of a finite state machine and added to it turing machine power through the use of actions on the arcs. It seems to be necessary to provide some representational means that is relatively easy to think about as a base and then contemplate how these simpler base forms can be modified to provide for the range of actual facts of natural language.

Moving to today's emphasis, we see increased interest in psychological reality. An example of this work is the thesis of Mitch Marcus, which attempts to deal with constraints imposed by human performance, as well as constraints of a more universal nature recently characterized by linguists. This model has been extended further by Bob Berwick to serve as the basis for a learning model. Another recent trend that causes me to smile a little is the resurgence of interest in context free grammars. I think back to Lyons' book on theoretical linguistics where context free grammar is chastised as was the custom, due to its inability to insightfully characterize subject-verb agreement, discontinuous constituents, and other things thought inappropriate for context free grammars. The fact that a context free grammar can always characterize any finite segment of the language was not a popular notion in the early days. Now we find increasing concern with efficiency arguments, and also due to the increasing emphasis in trying to find the simplest possible grammatical formalism to describe the facts of language, a vigorous effort to provide context free systems that provide a great deal of coverage. In the earlier days, the necessity of introducing additional non-terminals to deal with problems such as subject-verb agreement was seen as a definite disadvantage, but today such criticisms are hard to find. An additional trend that is interesting to observe is the current emphasis on ill-formed sentences which are now recognized as valid exemplars of the language and with which we must deal in a variety of computational applications. Thus, there has been attention focused on relaxation techniques and the

ability to parse limited phrases within discourse structures that may be ill-formed.

In the early days of the ACL, I believe that computation was seen mainly as a tool used to represent algorithms and provide for their execution. Now there is a much different emphasis on computation. Computing is seen as a metaphor, and as an important means to model various linguistic phenomena, as well as more broadly cognitive phenomena. This is an important trend, and is due in part to the emphasis in cognitive science on representational issues. When we must deal representations explicitly, then the branch of knowledge that provides the most help is computer science, and this fact is becoming much more widely appreciated, even by those workers who are not focused primarily on computing. This is a healthy trend, I believe, but we need also to be aware of the possibility of introducing biases and constraints on our thinking dictated by our current understanding and view of computation. Since our view of computation is in turn conditioned very substantially by the actual computing technology that is present at any given time, it is well to be very cautious in attributing basic understanding of these representations. A particular case in point is the emphasis, quite popular today, on parallelism. When we were used to thinking of computation solely in terms of single-sequence Von Neumann machines, then parallelism did not enjoy a prominent place in our models. Now that it is possible technologically to implement a great deal of parallelism, one can even discern more of a move to breadth first rather than depth first analyses. It seems clear that we are still very much the children of the technology that surrounds us.

I want to turn my attention now to a discussion of the development of speech processing technology, in particular, text-to-speech conversion and speech recognition, during the last twenty years. Speech has been studied over many decades, but its secrets have been revealed at a very slow pace. Despite the substantial in fusion of money into the study of speech recognition in the seventies, there still seems to be a natural gestation period for achieving new understanding of such complicated phenomena. Nevertheless, during these last twenty years, a great deal of useful speech processing capability has been achieved. Not only has there been much achievement, but these results have achieved great prominence through their coupling with modern technology. The outstanding example in speech synthesis technology has been of course the Texas Instruments Speak and Spell which demonstrated for the first time that acceptable use of synthetic speech could be achieved for a very modest price. Currently, there are at least 20 different integrated circuits, either already fabricated or under development, for speech synthesis. So a huge change has taken place. It is possible today to produce highly intelligible synthetic speech from text, using a variety of techniques in computational linguistics, including morphological analysis, letter-to-sound rules, lexical stress, syntactic parsing, and prosodic analysis. While

this speech can be highly intelligible, it is certainly not very natural yet. This reflects in part the fact we have been able to determine sufficient correlates for the percepts that we want to convey, but that we have thus far been unable to characterize the redundant interaction of a large variety of correlates that lead to integrated percepts in natural speech. Even such simple distinctions as the voiced/unvoiced contrast are marked by more than a dozen different correlates. We simply don't know, even after all these years, how these different correlates are interrelated as a function of the local context. The current disposition would lead one to hope that this interaction is deterministic in nature, but I suppose there is still some segment of the research community that has no such hopes. When the redundant interplay of correlates is properly understood, I believe this will herald a new improvement in our understanding needed for high performance speech recognition systems. Nevertheless, it is important to emphasize that during these twenty years, commercially acceptable text-to-speech systems have become viable, as well as many other speech synthesis systems utilizing parametric storage or waveform coding techniques of some sort.

Speech recognition has undergone a lot of change during this period also. The systems that are available in the marketplace are still based exclusively on template matching techniques, which probably have little or nothing to do with the intrinsic nature of speech and language. That is to say, they use some form of informationally-reduced representation of the input speech waveform and then contrive to match this representation against a set of stored templates. Various techniques have been introduced to improve the accuracy of this matching procedure by allowing for modifications of the input representation or the stored templates. For example, the use of dynamic programming to facilitate matching has been very popular, and for good reason, since its use has led to improvements in accuracy of between 20 and 30 percent. Nevertheless, I believe that the use of dynamic programming will not remain over the long pull and that more phonetically and linguistically based techniques will have to be used. This prediction is predicated, of course, on the need for a huge amount of improved understanding of language in all of its various representations and I feel that there is need for an incredibly large amount of new data to be acquired before we can hope to make substantial progress on these issues. Certainly an important contribution of computational linguistics is the provision of instrumental means to acquire data. In my view, the study of both speech synthesis and speech recognition has been hampered over the years in large part due to the sheer lack of insufficient data on which to base models and theories. While we would still like to have more computational power than we have, at present, we are able to provide highly capable interactive research environments for exploring new areas. The fact that there is none too much of these computational resources is supported by the fact that the speech

recognition group at IBM is, I believe, the
largest user of 370/168 time at Yorktown Heights.

and behavior.

An interesting aspect of the study of speech
recognition is that there is still no agreement
among researchers as to the best approach. Thus,
we see techniques based on statistical decoding,
those based on template matching using dynamic
programming, and those that are much more phonetic
and linguistic in nature. I believe that the
notion, at one time prevalent during the
seventies, that the speech waveform could often be
ignored in favor of constraints supplied by
syntax, semantics, or pragmatics is no longer held
and there is an increasing view that one should
try to extract as much information as possible
from the speech waveform. Indeed, word boundary
effects and manifestations at the phonetic level
of high level syntactic and semantic constraints
are being discovered continually as research in
speech production and perception continues. For
all of our research into speech recognition, we
are still a long ways away from approximating
human speech perception capability. We really
have no idea as to how human listeners are able to
adapt to a large variety of speakers and a large
variety of communication environments. We have no
idea how humans manage to reject noise in the
background, and very little understanding as to
the interplay of the various constraint domains
that are active. Within the last five years,
however we have seen an increasing level of
cooperation between linguists, psycholinguists
and computational linguists on these matters and
I believe that the depth of understanding in
psycholinguistics is now at a level where it can
be tentatively exploited by computational
linguists for models of speech perception.

Over these twenty years, we have seen
computational linguistics grow from a relatively
esoteric academic discipline to a robust
commercial enterprise. Certainly the need within
industry for man-machine interaction is very
strong and many computer companies are hiring
computational linguists to provide for natural
language access to data bases, speech control of
instruments, and audio announcements of all sorts.
There is a need to get newly developed ideas into
practice, and as a result of that experience,
provide feedback to the models that computational
linguists create. There is a tension, I believe,
between, on the one hand, the need to be far
reaching in our research programs vs. the need
for short-term payoff in industrial practice. It
is important that workers in the field seek to
influence those that control resources to maintain
a healthy balance between these two influences.
For example, the relatively new interest in
studying discourse structure is a difficult, but
important area for long range research and it
deserves encouragement, despite the fact that
there are large areas of ignorance and the need
for extended fundamental research. One can hope
however, that the demonstrated achievement of
computational linguistics over the last twenty
years will provide a base upon which society will
be willing to continue to support us to further
explore the large unknowns in language competence