

Current Research in the Development of a Spoken Language Understanding System using PARSEC*

Carla B. Zoltowski
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

February 28, 1991

1 Introduction

We are developing a spoken language system which would more effectively merge natural language and speech recognition technology by using a more flexible parsing strategy and utilizing prosody, the suprasegmental information in speech such as stress, rhythm, and intonation. There is a considerable amount of evidence which indicates that prosodic information impacts human speech perception at many different levels [5]. Therefore, it is generally agreed that spoken language systems would benefit from its addition to the traditional knowledge sources such as acoustic-phonetic, syntactic, and semantic information. A recent and novel approach to incorporating prosodic information, specifically the relative duration of phonetic segments, was developed by Patti Price and John Bear [1, 4]. They have developed an algorithm for computing break indices using a hidden Markov model, and have modified the context-free grammar rules to incorporate links between non-terminals which corresponded to the break indices. Although incorporation of this information reduced the number of possible parses, the processing time increased because of the addition of the link nodes in the grammar.

2 Constraint Dependency Grammar

Instead of using context-free grammars, we are using a natural language framework based on the

Constraint Dependency Grammar (CDG) formalism developed by Maruyama [3]. This framework allows us to handle prosodic information quite easily. Rather than coordinating lexical, syntactic, semantic, and contextual modules to develop the meaning of a sentence, we apply sets of lexical, syntactic, prosodic, semantic, and pragmatic rules to a packed structure containing a developing picture of the structure and meaning of a sentence. The CDG grammar has a weak generative capacity which is strictly greater than that of context-free grammars and has the added advantage of benefiting significantly from a parallel architecture [2]. PARSEC is our system based on the CDG formalism.

To develop a syntactic and semantic analysis using this framework, a network of the words for a given sentence is constructed. Each word is given some number indicating its position relative to the other words in the sentence. Once a word is entered in the network, the system assigns all of the possible *roles* the words can have by applying the lexical constraints (which specify legal word categories) and allowing the word to modify all the remaining words in the sentence or no words at all. Each of the arcs in the network has associated with it a matrix whose row and column indices are the roles that the words can play in the sentence. Initially, all entries in the matrices are set to one, indicating that there is nothing about one word's function which prohibits another word's right to fill a certain role in the sentence. Once the network is constructed, additional constraints are introduced to limit the role of each word in the sentence to a single function. In a spoken language system which may contain several possible candidates for each word, constraints would also

*Parallel Architecture Sentence Constrainer

provide feedback about impossible word candidates.

We have been able to incorporate the durational information from Bear and Price quite easily into our framework. An advantage of our approach is that the prosodic information is added as constraints instead of incorporating it into a parsing grammar. Because CDG is more expressive than context-free grammars, we can produce prosodic rules that are more expressive than Bear and Price are able to provide by augmenting context-free grammars. Also by formulating prosodic rules as constraints, we avoid the need to clutter our rules with nonterminals required by context-free grammars when they are augmented to handle prosody. Assuming $O(n^4/\log(n))$ processors, the cost of applying each constraint is $O(\log(n))$ [2]. Whenever we apply a constraint to the network, our processing time is incremented by this amount. In contrast, Bear and Price, by doubling the size of the grammar are multiplying the processing time by a factor of 8 when no prosodic information is available (assuming $(2n)^3 = 8n^3$ time).

3 Current Research

Our current research effort consists of the development of algorithms for extracting the prosodic information from the speech signal and incorporation of this information into the PARSEC framework. In addition, we will be working to interface PARSEC with the speech recognition system being developed at Purdue by Mitchell and Jamieson.

We have selected a corpus of 14 syntactically ambiguous sentences for our initial experimentation. We have predicted what prosodic features humans use to disambiguate the sentences and are attempting to develop algorithms to extract those features from the speech. We are hoping to build upon those algorithms presented in [1, 4, 5]. Initially we are using a professional speaker trained in prosodics in our experiments, but eventually we will test our results with an untrained speaker.

Although our current system allows multiple word candidates, it assumes that each of the possible words begin and end at the same time. It currently does not allow for non-aligned word boundaries. In addition, the output of the speech recognition system which we will be utilizing will

consist of the most likely sequence of phonemes for a given utterance, so additional work will be required to extract the most likely word candidates for use in our system.

4 Conclusion

The CDG formalism provides a very promising framework for our spoken language system. We believe its flexibility will allow it to overcome many of the limitations imposed by natural language systems developed primarily for text-based applications, such as repeated words and false starts of phrases. In addition, we believe that prosody will help to resolve the ambiguity introduced by the speech recognition system which is not present in text-based systems.

5 Acknowledgement

This research was supported in part by NSF IRI-9011179 under the guidance of Profs. Mary P. Harper and Leah H. Jamieson.

References

- [1] J. Bear and P. Price. Prosody, syntax, and parsing. In *Proceedings of the 28th annual ACL*, 1990.
- [2] R. Helzerman and M.P. Harper. Parsec: An architecture for parallel parsing of constraint dependency grammars. In *Submitted to The Proceedings of the 29th Annual Meeting of ACL*, June 1991.
- [3] H. Maruyama. Constraint dependency grammar. Technical Report #RT0044, IBM, Tokyo, Japan, 1990.
- [4] P. Price, C. Wightman, M. Ostendorf, and J. Bear. The use of relative duration in syntactic disambiguation. In *Proceedings of ICSLP*, 1990.
- [5] A. Waibel. *Prosody and Speech Recognition*. Morgan Kaufmann Publishers, Los Altos, CA, 1988.