

# Prosodic Aids to Syntactic and Semantic Analysis of Spoken English

Chris Rowles and Xiuming Huang

AI Systems Section

Australia and Overseas Telecommunications Corporation

Telecommunications Research Laboratories

PO Box 249, Clayton, Victoria, 3168, Australia

Internet: c.rowles@trf.oz.au

## ABSTRACT

Prosody can be useful in resolving certain lexical and structural ambiguities in spoken English. In this paper we present some results of employing two types of prosodic information, namely pitch and pause, to assist syntactic and semantic analysis during parsing.

## 1. INTRODUCTION

In attempting to merge speech recognition and natural language understanding to produce a system capable of understanding spoken dialogues, we are confronted with a range of problems not found in text processing.

Spoken language conversations are typically more terse, less grammatically correct, less well-structured and more ambiguous than text (Brown & Yule 1983). Additionally, speech recognition systems that attempt to extract words from speech typically produce word insertion, deletion or substitution errors due to incorrect recognition and segmentation.

The motivation for our work is to combine speech recognition and natural language understanding (NLU) techniques to produce a system which can, in some sense, understand the intent of a speaker in telephone-based, information seeking dialogues. As a result, we are interested in NLU to improve the semantic recognition accuracy of such a system, but since we do not have explicit utterance segmentation and structural information, such as punctuation in text, we have explored the use of prosody.

Intonation can be useful in understanding dialogue structure (c.f. Hirschberg & Pierrehumbert 1986), but parsing can also be assisted. (Briscoe & Boguraev 1984) suggests that if prosodic structure could be derived for the noun compound *Boron epoxy rocket motor chambers*, then their parser LEXICAT could reduce the fourteen licit

morphosyntactic interpretations to one correct analysis without error (p. 262). (Steedman 1990) explores taking advantage of intonational structure in spoken sentence understanding in the combinatory categorial grammar formalism. (Bear & Price 1990) discusses integrating prosody and syntax in parsing spoken English, relative duration of phonetic segments being the one aspect of prosody examined.

Compared with the efforts expended on syntactic/semantic disambiguation mechanisms, prosody is still an under-exploited area. No work has yet been carried out which treats prosody at the same level as syntax, semantics, and pragmatics, even though evidence shows that prosody is as important as the other means in human understanding of utterances (see, for example, experiments reported in (Price et al 1989)). (Scott & Cutler 1984) noticed that listeners can successfully identify the intended meaning of ambiguous sentences even in the absence of a disambiguating context, and suggested that speakers can exploit acoustic features to highlight the distinction that is to be conveyed to the listener (p. 450).

Our current work incorporates certain prosodic information into the process of parsing, combining syntax, semantics, pragmatics and prosody for disambiguation<sup>1</sup>. The context of the work is an electronic directory assistance system (Rowles et al 1990). In the following sections, an overview of the system is first given (Section 2). Then the parser is described in Section 3. Section 4 discusses how prosody can be employed in helping resolve ambiguity involved in process-

---

1. Another possible acoustic source to help disambiguation is "segmental phonology", the application of certain phonological assimilation and elision rules (Scott & Cutler 1984). The current work makes no attempt at this aspect.

ing fixed expressions, prepositional phrase attachment (PP attachment), and coordinate constructions. Section 5 shows the implementation of the parser.

## 2. SYSTEM OVERVIEW

Our work is aimed at the construction of a prototype system for the understanding of spoken requests to an electronic directory assistance service, such as finding the phone number and address of a local business that offers particular services.

Our immediate work does not concentrate on speech recognition (SR) or lexical access. Instead, we assume that a future speech recognition system performs phoneme recognition and uses linguistic information during word recognition. Recognition is supplemented by a prosodic feature extractor, which produces features synchronized to the word string output by the SR.

The output of the recognizer is passed to a sentence-level parser. Here "sentence" really means a conversational move, that is, a contiguous utterance of words constructed so as to convey a proposition.

Parses of conversational moves are passed to a dialogue analyzer that segments the dialogue into contextually-consistent sub-dialogues (i.e. exchanges) and interpret speaker requests in terms of available system functions. A dialogue manager manages interaction with the speaker and retrieves database information.

## 3. PROSODY EXTRACTION

As the input to the parser is spoken language, it lacks the segmentation apparent in text. Within a move, there is no punctuation to hint at internal grammatical structure. In addition, as complete sentences are frequently reduced to phrases, ellipsis etc. during a dialogue, the Parser cannot use syntax alone for segmentation.

Although intonation reflects deeper issues, such as a speakers' intended interpretation, it provides the surface structure for spoken language. Intonation is inherently supra-segmental, but it is also useful for segmentation purposes where other information is unavailable. Thus, intonation can be used to provide initial segmentation via a pre-processor for the parser.

Although there are many prosodic features that are potentially useful in the understanding of spoken English, pitch and pause information have received the most attention due to ease of measurement and their relative importance (Cruttenden 1986, pp 3 & 36). Our efforts to date use only these two feature types.

We extract pitch and pause information from speech using specifically designed hardware with some software post-processing. The hardware performs frequency to amplitude transformation and filtering to produce an approximate pitch contour with pauses.

The post-processing samples the pitch contour, determines the pitch range and classifies the instantaneous pitch into high, medium and low categories within that range. This is similar to that used in (Hirschberg & Pierrehumbert 1986). Pauses are classed as short (less than 250ms), long (between 250ms and 800ms) or extended (greater than 800ms). These times were empirically derived from spoken information seeking dialogues conducted over a telephone to human operators. Short pauses signify strong turn-holding behaviour, long pauses signify weaker turn-holding behaviour and extended pauses signify turn passing or exchange completion (Vonwiller 1991). These interpretations can vary with certain pitch movements, however. Unvoiced sounds are distinguished from pauses by subsequent synchronisation of prosodic features with the word stream by post-processing.

A parser pre-processor then takes the SR word string, pitch markers and pauses, annotating the word string with pitch markers (low marked as " ~ ", medium " - " and high " ^ ") and pauses (short " \* " and long " \*\* "). The markers are synchronised with words or syllables. The pre-processor uses the pitch and pause markers to segment the word string into intonationally-consistent groups, such as tone groups (boundaries marked as " < " and " > ") and moves (/). A tone group is a group of words whose intonational structure indicates that they form a major structural component of the speech, which is commonly also a major syntactic grouping (Cruttenden 1986, pp. 75 - 80). Short conversational moves often correspond to tone groups, while longer moves may consist of several tone groups. With cue words for example, the cue forms its own tone group.

Pauses usually occur at points of low transitional probability and often mark phrase boundaries (Cruttenden 1986). In general, although pitch plays an important part, long pauses, indicate tone group and move boundaries, and short pauses indicate tone group boundaries. Exchange boundary markers are dealt with in the dialogue manager (not covered here). Pitch movements indicate turn-holding behaviour, topic changes, move completion and information contrastiveness (Cooper & Sorensen 1977; Vonwiller 1991).

The pre-processor also locates fixed expressions, so that during the parsing nondeterminism can be reduced. A problem here is that a cluster of words may be ambiguous in terms of whether they form a fixed expression or not. "Look after", for example, means "take care of" in "*Mary helped John to look after his kids*", whereas "look" and "after" have separate meaning in "*I'll look after you do so*". The pre-processor makes use of tone group information to help resolve the fixed expression ambiguity. A more detailed discussion is given in section 5.2.

#### 4. THE PARSER

Once the input is segmented, moves annotated with prosody are input to the parser. The parser deals with one move at a time.

In general, the intonational structure of a sentence and its syntactic structure coincide (Cruttenden 1986). Thus, prosodic segmentation avoids having the Parser try to extract moves from unsegmented word strings based solely on syntax. It also reduces the computational complexity in comparing syntactic and prosodic word groupings. There is a complication, however, in that tone group boundaries and move boundaries may not align exactly. This is not frequent, and is not present in the material used here. Intonation is used to limit the range of syntactic possibilities and the parser will align tone group and move syntactic boundaries at a later stage.

By integrating syntax and semantics, the Parser is capable of resolving most of the ambiguous structures it encounters in parsing written English sentences, such as coordinate conjunctions, PP attachments, and lexical ambiguity (Huang 1988). Migrating the Parser from written to spoken English is our current focus.

Moves input to the Parser are unlikely to be well-formed sentences, as people do not always speak grammatically, or due to the SR's inability to accurately recognise the actual words spoken. The parser first assumes that the input move is lexically correct and tries to obtain a parse for it, employing syntactic and semantic relaxation techniques for handling ill-formed sentences (Huang 1988). If no acceptable analysis is produced, the parser asks the SR to provide the next alternative word string.

Exchanges between the parser and the SR are needed for handling situations where an ill-formed utterance gets further distorted by the SR. In these cases other knowledge sources such as pragmatics, dialogue analysis, and dialogue management must be used to find the most likely interpretation for the input string. We use pragmatics and knowledge of dialogue structure to find the semantic links between separate conversational moves by either participant and resolve indirectness such as pronouns, deictic expressions and brief responses to the other speaker [for more details, see (Rowles, 1989)]. By determining the dialogue purpose of utterances and their domain context, it is then possible to correct some of the insertion and mis-recognised word errors from the SR and determine the communicative intent of the speaker. The dialogue manager queries the speaker if sentences cannot be analysed at the pragmatic stage.

The output of the parser is a parse tree that contains syntactic, semantic and prosodic features. Most ambiguity is removed in the parse tree, though some is left for later resolution, such as definite and anaphoric references, whose resolution normally requires inter-move inferences.

The parser also detects cue words in its input using prosody. Cue words, such as "now" in "*Now, I want to ...*", are words whose meta-function in determining the structure of dialogues overrides their semantic roles (Reichman 1985). Cue words and phrases are prosodically distinct due to their high pitch and pause separation from tone groups that convey most of the propositional content (Hirschberg & Litman 1987). While relatively unimportant semantically, cue words are very important in dialogue analysis due to their ability to indicate segmentation and the linkage of the dialogue components.

## 5. PROSODY AND DISAMBIGUATION

During parsing prosodic information is used to help disambiguate certain structures which cannot be disambiguated syntactically/semantically, or whose processing demands extra efforts, if no such prosodic information is available. In general, prosody includes pitch, loudness, duration (of words, morphemes and pauses) and rhythm. While all of these are important cues, we are currently focussing on pitch and pauses as these are easily extracted from the waveform and offer useful disambiguation during parsing and segmentation in dialogue analysis. Subsequent work will include the other features, and further refinement of the use of pitch and pause. At present, for example, we do not consider the length of pauses internal to tone groups, although this may be significant.

The prosodic markers are used by the parser as additional pre-conditions for grammatical rules, discriminating between possible grammatical constructions via consistent intonational structures.

### 5.1 HOMOGRAPHS

Even when using prosody, homographs are a problem for parsers, although a system recognising words from phonemes can make the problem a simpler. The word sense of "bank" in "*John went to the bank*" must be determined from semantics as the sense is not dependent upon vocalisation, but the difference between the homograph "content" in "*contents of a book*" and "*happy and content*" can be determined through differing syllabic stress and resultant different phonemes. Thus, different homographs can be detected during lexical access in the SR independently of the Parser.

### 5.2 FIXED EXPRESSIONS

As is mentioned in subsection 4.1, when the pre-processor tries to locate fixed expressions, it may face multiple choices. Some fixed expressions are obligatory, i.e., they form single semantic units, for instance "look forward to" often means "expect to feel pleasure in (something about to happen)"<sup>2</sup>. Some other strings may or

---

2. Longman Dictionary of Contemporary English, 1978.

may not form single semantic units, depending on the context. "Look after" and "win over" are two examples. Without prosodic information, the pre-processor has to make a choice blindly, e.g. treating all potential fixed expressions as such and on backtracking dissolve them into separate words. This adds to the nondeterminism of the parsing. As prosodic information becomes available, the nondeterminism is avoided.

In the system's fixed expression lexicon, we have entries such as "fix\_e([gave, up], gave\_up)". The pre-processor contains a rule to the following effect, which conjoins two (or more) words into one fixed expression only when there is no pause following the first word:

```
match_fix_e([FirstW, SecondW|RestW], [FixedE|MoreW]):-
    no_pause_in_between(FirstW, SecondW),
    fix_e([FirstW, SecondW], FixedE),
    Match_fix_e(RestW, MoreW).
```

This rule produces the following segmentations:

(5.1a) <~He ~gave> \*^up to ^two hundred dollars> \*^-to the ^charity>\*\*//

(5.1b) <~He ^gave ^up> \*^two hundred dollars> \*^-for damage compensation>\*\*//.

In (5.1a), *gave* and *up* to are treated as belonging to two separate tone groups, whereas in (5.1b) *gave up* is marked as one tone group. The pre-processor checking its fixed expression dictionary will therefore convert *up to* in (5.1a) to *up\_to*, and *gave up* in (5.1b) to *gave\_up*.

### 5.3 PP ATTACHMENT

(Steedman 1990 & Cruttenden 1986) observed that intonational structure is strongly constrained by meaning. For example, an intonation imposing bracketings like the following is not allowed:

(5.2) <Three cats> <in ten prefer corduroy>//

Conversely, the actual contour detected for the input can be significant in helping decide the segmentation and resolving PP attachment. In the following sentence, f.g.,

(5.3) <I would like> <information on her arrival> ["on her arrival" attached to "information"]

(5.4) <I would like> <information> \*\* <on her arrival> ["on her arrival" attached to "like"]

the pause after "information" in (5.4), but not in (5.3), breaks the bracketed phrase in (5.3) into two separate tone groups with different attachments.

In a clash between prosodic constraints and syntactic/semantic constraints, the latter takes precedence over the former. For instance, in:

(5.5) <I would like> <information> \*\* <on some panel beaters in my area>.

although the intonation does not suggest attachment of the PP to "information", since the semantics constraints exclude attachment to "like" meaning "choose to have" ("On panel beaters [as a location or time] I like information" does not rate as a good interpretation), it is attached to "information" anyway (which satisfies the syntactic/semantic constraints).

#### 5.4 COORDINATE CONSTRUCTIONS

Coordinate constructions can be highly ambiguous, and are handled by rules such as:

Np --> det(Det), adj(Adj),  
 /\* check if a pause follows the adjective \*/  
 {check\_pause (Flag)}, noun (Noun),  
 {construct\_np(Det, Adj, Noun, NP)},  
 conjunction(NP, Flag, FinalNP).

In the conjunction rule, if two noun phrases are joined, we check for any pauses to see if the adjective modifying the first noun should be copied to allow it to modify the second noun. Similarly, we check for a pause preceding the conjunction to decide if we should copy the post modifier of the second noun to the first noun phrase. For instance, the text-form phrase:

(5.6) old men and women in glasses

can produce three possible interpretations:

[old men (in glasses)] and [(old) women in glasses] (5.6a)

[old men] and [women in glasses] (5.6b)

[old men (in glasses)] and [women in glasses] (5.6c).

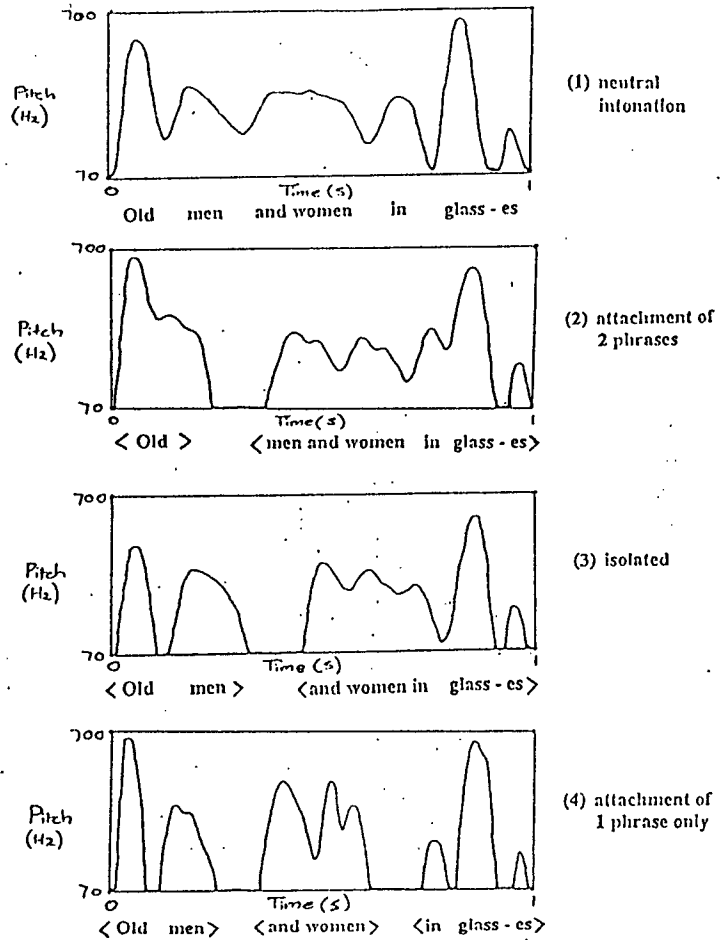


Figure 1.

Figure 1 shows some measured pitch contours for utterances of phrase (5.6) with an attempt by the speaker to provide the interpretations (a) through (c). Note that the contour is smoothed by the hardware pitch extraction. Pauses and unvoiced sounds are distinguished in the software post-processor.

In all waveforms "old" and "glasses" have high pitch. In (5.6a), a short pause follows "old", indicating that "old" modifies "men and women in glasses" as a sub-phrase. This is in contrast to (5.6b) where the short pause appears after "men" indicating "old men" as one conjunct and "women in glasses" as the other. Notice also that duration of "men" in (5.6b) is longer than in (5.6a). In (5.6c) we have two major pauses, a shorter one after "men" and a longer one after "women". Using this variation in pause locations,

the parser produces the correct interpretation (i.e. the speaker's intended interpretation) for sentences (5.6a-c).

## 6. IMPLEMENTATION

Prosodic information, currently the pitch contour and pauses, are extracted by hardware and software. The hardware detects pitch and pauses from the speech waveform, while the software determines the duration of pauses, categorises pitch movements and synchronises these to the sequence of lexical tokens output from a hypothetical word recogniser. The parser is written in the Definite Clause Grammars formalism (Pereira et al. 1980) and runs under BIMProlog on a SPARCstation 1. The pitch and pause extractor as described here is also complete.

To illustrate the function of the prosodic feature extractor and the Parser pre-processor, the following sentence was uttered and its pitch contour analysed:

"yes i'd like information on some panel beaters"

Prosodic feature extraction produced:

```
** ^yes ** ^i'd ^like * -information on some ^panel
beaters **//
```

The Parser pre-processor then segments the input (in terms of moves and tone groups) for the Parser, resulting in:

```
**< ^yes> **// < ^i'd ^like> * <-information on some
^panel beaters> **//
```

The actual output of the pre-processor is in two parts, one an indexed string of lexical items plus prosodic information, the other a string of tone groups indicating their start and end points:

```
[** ^yes, 1] [**// ^i, 2] [would, 3] [^like, 4] [* -infor-
mation, 5] [on, 6] [some, 7] [^panel_ beaters, 8]
[**//, 9]
```

```
<1,1> <2, 4> <5, 8> <9,9>
```

We use a set of sentences<sup>3</sup>, all beginning with "Before the King/feature races", but with different intonation to provide different interpretations, to illustrate how syntax, semantics and

3. Adapted from (Briscoe & Boguraev 1984).

prosody are used for disambiguation:

```
(6.1) <~ Before the -King ^races>* <-his
^horse> <is -usually ^groomed>**//.
```

```
(6.2) <~Before the -King> * <-races his
^horse> **<it's -usually ^groomed>**//.
```

```
(6.3) <~Before the ^feature ~races> * <-his
^horse is -usually ^groomed>**//.
```

The syntactic ambiguity of "before" (preposition in 6.3 and subordinate conjunction in 6.1 and 6.2) is solved by semantic checking: "race" as a verb requires an animate subject, which "the King" satisfies, but not "the feature"; "race" as a noun can normally be modified by other nouns such as "feature", but not "King"<sup>4</sup>. However, when prosody information is not used the time needed for parsing the three sentences varies tremendously, due to the top-down, depth-first nature of the parser. (6.3) took 2.05 seconds to parse, whereas (6.1) took 9.34 seconds, and (6.2), 41.78 seconds. The explanation lies in that on seeing the word "before" the parser made an assumption that it was a preposition (correct for 6.3), and took the "wrong" path before backtracking to find that it really was a conjunction (for 6.1 and 6.2). Changing the order of rules would not help here: if the first assumption treats "before" as a conjunction, then parsing of (6.3) would have been slowed down.

We made one change to the grammar so that it takes into account the pitch information accompanying the word "races", to see if improvement can be made. The parser states that a noun-noun string can form a compound noun group only when the last noun has a low pitch. That is, *the feature ~races* forms a legitimate noun phrase, while *the King -races* and *the King ^races* do not. This is in accordance with one of the best known English stress rules, the "Compound Stress Rule" (Chomsky and Halle 1968), which asserts that the first lexically stressed syllable in a constituent has the primary stress if the constituent is a compound construction forming an adjective, verb, or noun.

4. It is very difficult, though, to give a clear cut as to what kind of nouns can function as noun modifiers. King races may be a perfect noun group in certain context.

We then added the pause information in the parser along similar lines. The following is a simplified version of the VP grammar to illustrate the parsing mechanism:

```

/* Noun phrase rule.
"Mods" can be a string of adjectives or nouns:
major (races), feature (races), etc.*/
Np --> Det, Mods,HeadNoun.
/* Head noun is preferred to be low-pitched.*/
HeadNoun --> [Noun], {low_pitched(Noun)}.

/* Verb phrase rule 1.*/
Vp --> V_intr.

/* Verb phrase rule 2. Some semantic checking
is carried out after a transitive verb and a
noun phrase is found.*/
Vp --> V_tr, Np, {match(V_tr,Np)}.
/* If a verb is found which might be used as in-
transitive, check if there is a pause following it.*/
V_intr --> [Verb], {is_intransitive(Verb)},
Pause.
/* Otherwise see if the verb can be used as
transitive.*/
V_tr --> [Verb], {is_transitive(Verb)}.
/* This succeeds if a pause is detected. */
Pause --> [pause].

```

The pause information following "races" in sentences(6.1) and (6.2) thus helps the parser to decide if "races" is transitive or intransitive, again reducing nondeterminism. The above rules specify only the preferred patterns, not absolute constraints. If they cannot be satisfied, e.g. when there is no pause detected after a verb which is intransitive, the string is accepted anyway.

The parse times for sentences (6.1) to (6.3) with and without prosodic rules in the parser are given in the Table 6.1.

	Without Prosody	With Prosody
(6.1)	9.34	1.23
(6.2)	41.78	8.69
(6.3)	2.05	1.27

**Table 6.1** Parsing Times for the "races" sentence (in seconds).

Table 6.2 shows how the parser performed on the following sentences:

(6.4) ^I'll look\* ^after the -boy ~comes\*\*//

(6.5) ^He ^gave\* ^up to ^two ^hundred dollars to the -charity\*\*//

(6.6) ^Now\* -I want -some -information on ^panel ^beaters -in ~Clayton\*\*//

	Without Prosody	With Prosody
(6.4)	6.59	1.19
(6.5)	41.38	2.49
(6.6)	2.15	2.55

**Table 6.2** Parsing Times for sentences (6.4) to (6.6) (in seconds).

While (6.6) is slower with prosodic annotation, the parser correctly recognises "now" as a cue word rather than as an adverb.

## 7. DISCUSSION

We have shown that by integrating prosody with syntax and semantics in a natural language parser we can improve parser performance. In spoken language, prosody is used to isolate sentences at the parser's input and again to determine the syntactic structure of sentences by seeking structures that are intonationally and syntactically consistent.

The work described here is in progress. The prosodic features with which sentences have been annotated are the output of our feature extractor, but synchronisation is by hand as we do not have a speech recognition system. As shown by the "old men ..." example, the system is capable of accurately producing correct interpretations, but as yet, no formal experiments using data extracted from ordinary telephone conversations and human comparisons have been performed. The aim has been to investigate the potential for the use of prosody in parsers intended for use in speech understanding systems.

(Bear & Price 1990) modified the grammar they use to change all the rules of the form A -> B C to the form A -> B Link C, and add constraints to the rules application in terms of the value of the "breaking indices" based on relative duration of phonetic segments. For instance the rule VP -> V Link PP applies only when the value of the link is either 0 or 1, indicating a close coupling of neighbouring words. Duration is thus tak-

en into consideration in deciding the structure of the input. In our work, pitch contour and pause are used instead, achieving a similar result.

The principle of preference semantics allows the straightforward integration of prosody into parsing rules and a consistent representation of prosody and syntax. Such integration may have been more of a problem if the basic parsing approach had been different. Also relevant is the choice of English, as the integration may not carry across to other languages.

Future research aims at a more thorough treatment of prosody. Research currently underway, is also focussing on the use of prosody and dialogue knowledge for dialogue analysis and turn management.

#### ACKNOWLEDGEMENTS

The permission of the Director, Research, AOTC to publish the above paper is hereby acknowledged. The authors have benefited from discussions with Robin King, Peter Sefton, Julie Vonwiller and Christian Matthiessen, Sydney University, and Muriel de Beler, Telecommunication Research Laboratories, who are involved in further work on this project. The authors would also like to thank the anonymous reviewers for positive comments on paper improvements.

#### REFERENCES

Bear, J. & Price, P. J. (1990), Prosody, Syntax and Parsing. *28th Annual Meeting of the Assoc. for Computational Linguistics* (pp. 17-22).

Briscoe, E.J. & Boguraev, B.K. (1984), Control Structures and Theories of Interaction in Speech Understanding Systems. *22th Annual Meeting of the Assoc. for Computational Linguistics* (pp. 259-266)

Brown, G., & Yule, G., (1983), *Discourse Analysis*, Cambridge University Press.

Chomsky, N. & Halle, M. (1968), *The Sound Pattern of English*, (New York: Harper and Row).

Cooper, W.E. & Sorensen, J.M., (1977), Fundamental Frequency Contours at Syntactic Boundaries, *Journal of the Acoustical Society of America*, Vol. 62, No. 3, September.

Cruttenden, A., (1986), *Intonation*, Cam-

bridge University Press.

Hirschberg, J. & Litman, D., (1987), Now Let's Talk About Now: Identifying Cue Phrases Intonationally, *25th Annual Meeting of the Assoc. for Computational Linguistics*.

Hirschberg, J. & Pierrehumbert, J., The International Structure of Discourse, *24th Annual Meeting of the Assoc. for Computational Linguistics*, 1986.

Huang, X-M. (1988), Semantic Analysis in XTRA, An English - Chinese Machine Translation System, *Computers and Translation* 3, No.2. (pp. 101-120)

Pereira, F. & Warren, D. (1980), Definite Clause Grammars for Language Analysis - A Survey of the Formalism and A Comparison with Augmented Transition Networks. *Artificial Intelligence*, 13:231-278.

Price, P. J., Ostendorf, M. & Wightmen, C.W. (1989), Prosody and Parsing. *DARPA Workshop on Speech and Natural Language*, Cape Cod, October 1989 (pp.5-11).

Reichman, R. (1985), *Getting Computers to Talk Like You and Me*, (Cambridge: MIT Press).

Rowles, C.D. (1989), Recognizing User Intentions from Natural language Expressions, *First Australia-Japan Joint Symposium on Natural Language Processing*, (pp.157-166).

Rowles, C.D., Huang, X., and Aumann, G., (1990), Natural Language Understanding and Speech Recognition: Exploring the Connections, *Third Australian International Conference on Speech Science and Technology*, (pp. 374 - 382).

Steedman, M. (1990), Structure and Intonation in Spoken Language Understanding. *28th Annual Meeting of the Assoc. for Computational Linguistics* (pp. 9-16).

Scott, D.R & Cutler, A. (1984), Segmental Phonology and the Perception of Syntactic Structure, *Journal of Verbal Learning and Verbal Behavior* 23, (pp. 450-466).

Vonwiller, J. (1991), An Empirical Study of Some Features of Intonation, *Second Australia-Japan Natural Language Processing Symposium*, Japan, November, (pp 66-71).