Accent and Discourse Context: Assigning Pitch Accent in Synthetic Speech

Julia Hirschberg
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill NJ 07974
julia@research.att.com

Abstract

Identifying the regularities underlying speaker decisions to emphasize or de-emphasize an item intonationally has long been the subject of speculation and controversy. This paper describes a study of accent assignment based upon the analysis of natural recorded (read) speech. Results are being incorporated in **NewSpeak**, an interface to the Bell Laboratories Text-to-Speech System, which varies intonational features based upon syntactic structure and higher-level discourse information inferred from unrestricted text, in order to generate more natural synthetic speech. Implications of the work for the evaluation of discourse models, for automatic labeling of prosodic features, and for speech synthesis are discussed.

Introduction

The question of how speakers decide which items to emphasize intonationally and which to de-emphasize has been a popular but vexing problem in studies of intonation and discourse. While most researchers today accept Bolinger's pessimistic "Accent is Predictable (if you're a mind-reader)" (Bolinger 1972), mind-reading attempts continue, fueled in particular by the need to assign acceptable accentuation in speech synthesis. Although current systems generally use only simple syntactic information for this purpose, it is abundantly clear that 'higher-level' discourse information will have to be brought to bear as well, especially as applications requiring the synthesis of longer texts (as opposed to single words or sentences) become important. It also appears that pitch accent assignment, in turn, may provide a useful metric in the evaluation of current models of discourse. While the most objective criterion currently available for such evaluation has been the proper resolution of pronominal reference (Walker 1989), the ability of discourse models to support accent placement comparable to that observed in natural speech should provide a useful new evaluation criterion.

This paper describes the assignment of pitch accent in **NewSpeak**, an interface to the Bell Laboratories Text-to-Speech System (TTS) (Olive and Liberman 1985), which varies intonational features based upon syntactic structure and higher-level discourse information inferred from unrestricted text, in order to generate more natural speech. It compares the strategy of accent assignment initially implemented in NewSpeak with traditional strategies in speech synthesis – and with more sophisticated variants suggested by comparison with natural speech. First, previous work on the interpretation of pitch accent and on attempts to define rule systems for accent assignment are discussed, with an emphasis on speech synthesis efforts. Next, NewSpeak itself and its initial accent assignment algorithm are described. Then, the empirical data currently being used to refine this algorithm is examined, together with some preliminary results. Finally, future directions for the work are discussed.

Pitch Accent in Natural and Synthetic Speech

In natural speech, some words appear more intonationally prominent than others. We say that these words bear PITCH ACCENTS. While, in English, each word has a characteristic stress pattern, not every word is accented. Thus, LEXICAL STRESS must be distinguished from pitch accent. Although pitch accent is a perceptual phenomenon, words that hearers identify as accented tend to differ from their DEACCENTED versions (those not bearing pitch accents) in terms of some combination of pitch, duration, amplitude, and spectral characteristics. Accented words are usually identifiable in the FUNDAMENTAL FREQUENCY CONTOUR (f0) as local maxima or minima, aligned with the word's stressed syllable; deaccented items are **not** aligned with such pitch excursions. Words perceived as accented also tend to be somewhat longer and louder than deaccented words. The vowel in the stressed syllable of a deaccented word is often reduced from the full vowel of the accented version.

The role that accent plays in utterance interpretation is as yet not well understood. While attempts to explain the effect of pitch accent on the interpretation of particular syntactic, semantic, or pragmatic phenomena abound, and while some more general ac-

counts of accent meaning have been proposed either for particular word classes or for accent in general, the former accounts are too restricted and the latter too preliminary to provide much practical guidance for the construction of accent assignment algorithms.

Currently, most researchers recognize certain 'defaults' for accent placement, arising from some combination of word class, syntactic constituency, and surface position, which can be 'overridden' by phenomena described variously as CONTRASTIVENESS (Bolinger 1961), FOCUS (Jackendoff 1972; Rooth 1985), or INFORMATION STATUS (such as the GIVEN/NEW distinction) (Chafe 1976; Halliday and Hassan 1976; Clark and Clark 1977; Prince 1981). However, the difficulty of defining such phenomena and, consequently, of predicting when speakers will wish to convey them has made it difficult to incorporate higher-level discourse or intentional considerations into algorithms describing speakers' accent placement strategy. Nonetheless, whereas structural considerations were once thought to be primary in speakers' accent placement decisions (Quirk *et al.* 1964; Crystal 1975; Cruttenden 1986), now the role assigned to syntax in determining accent has shrunk markedly, as the role of discourse features and inferred speaker intention in determining accent location has grown. So the thrust of work on accent placement has been to seek solutions at the level of discourse context.

Much of the progress in both the interpretation and the location of pitch accent currently being made arises from experimental studies such as (Brown 1983) and analyses of large corpora of recorded speech such as (Altenberg 1987). However, applications of such results to speech synthesis are so far in their infancy. Most text-to-speech systems make accenting decisions based purely upon word class, with CLOSED-CLASS items (function words, such as articles, prepositions and pronouns) generally deaccented and OPEN-CLASS items (content words, such as nouns and verbs) accented. Such approaches tend to accent too many words in synthesis of longer stretches of text. And general inadequacies in part-of-speech tagging make it difficult to distinguish between items which are ostensibly closed-class but are commonly accented (e.g. verbal prepositions and particles) and those that are not.

Exceptions to the standard approach are found in message-to-speech systems, which synthesize speech from more abstract syntactic or semantic representations (Young and Fallside 1979; Danlos *et al.* 1986; Davis and Hirschberg 1988). And in one experimental text-to-speech system, (Silverman 1987) used a FIFO queue of roots of mentioned open-class items to determine whether subsequent items should be deaccented, clearing the queue at paragraph boundaries. But most experimental systems, and all commercial systems rely upon the simple closed-open class distinction, dealing with exceptions simply by permitting varying degrees of user ability to override incorrect system defaults.

Assigning Pitch Accent in NewSpeak

The development of NewSpeak as an interface to TTS was motivated by both theoretical and practical considerations. On the one hand, the improvement of speech analysis technology makes it feasible to test hypotheses about the relationship between intonational features and syntactic, semantic, and discourse characteristics of a text against large amounts of recorded speech. On the other hand, the deficiencies of current text-to-speech systems in synthesizing longer stretches of text, combined with the emergence of new applications for text-to-speech systems which require such synthesis, makes the improvement of strategies for assigning intonational features more pressing. Implemented in Quintus Prolog, NewSpeak currently uses syntactic information plus a simple discourse model to assign pitch accent type and placement, intonational phrasing, pausal duration, speaking rate, and phrasing in unrestricted text. Output from this interface in the form of text annotated with escape sequences which control intonational decisions is then passed to TTS for realization in speech. In this paper, NewSpeak's accent placement strategy will be described, together with preliminary results from current investigations which have improved that strategy by making more sophisticated use of discourse information.

NewSpeak originally assigned pitch accent using part-of-speech information provided by the Fidditch parser (Hindle 1989), a compound nominal parser based upon (Liberman and Sproat Forthcoming) (and implemented by Richard Sproat), and a partial model of the current text, based upon (Grosz and Sidner 1986)'s notion of the ATTENTIONAL STRUCTURE of discourse. A word's part-of-speech identified its default pitch accent: closed-class items were deaccented and open-class accented. Structural information also permited identification of possible nominal compounds, which NewSpeak passed to the compound stress assignment program for default (citation form) stress assignment (e.g. 'city *hall*' was assigned right stress and '*parking* lot', left stress). Output from this assignment was then passed to NewSpeak's discourse model.

The chief innovative aspect of accent assignment in NewSpeak was and is the use made of a discourse model. This model is based upon (Grosz and Sidner 1986)'s notion of a discourse's ATTENTIONAL STRUCTURE, and is implemented as a simple stack of sets of roots of open-class items mentioned in the text. In the (Grosz and Sidner 1986) model of discourse, discourse structure comprises three structures, a LINGUISTIC STRUCTURE, which is the text/speech itself; an ATTENTIONAL STRUCTURE, which includes information about the relative salience of objects, properties, relations, and intentions at any point in the discourse; and an INTENTIONAL STRUCTURE, which relates the intentions underlying the production of speech segments to one another. In NewSpeak, attentional structure is represented as a stack of FOCUS SPACES, each of which

comprises the set of roots of open-class words produced in the intonational phrase currently being constructed. The attentional state is updated — i.e., focus spaces are pushed or popped — by a simple algorithm which infers topic structure from orthographic cues, such as paragraphing and punctuation, and lexical cues, such as CUE PHRASES (words such as *now*, *well*, and *by the way*, which provide explicit information about the structure of a discourse); each cue phrase has associated with it a 'push' or 'pop' operation for this purpose (Litman and Hirschberg 1990). In NewSpeak's original discourse model, this stack was popped at each paragraph boundary. In addition to this model of LOCAL FOCUS (Sidner 1983; Grosz *et al.* 1983; Brennan *et al.* 1987), NewSpeak creates a GLOBAL FOCUS SPACE from the set of roots of open-class items mentioned in the first sentence of the text, assuming that that sentence can be treated as topic-defining. While the set of items in local focus is constantly subject to change, items in global focus remain so throughout the discourse. Note that no attempt has been made at this stage to model (Grosz and Sidner 1986)'s attentional structure entirely. In particular, no aspect of intentional structure is included, and objects, properties and relations are represented here by their roots rather than by some more abstract conceptual representation. Also, no positional distinction is made among items on in a focus space — or between focus spaces, except for the local-global distinction.

In NewSpeak's original accent assignment algorithm, items in either global or local focus were treated as GIVEN, representing OLD INFORMATION in the discourse (Halliday and Hassan 1976; Chafe 1976; Clark and Clark 1977; Prince 1981), which a speaker is entitled to believe is shared with his/her hearer. Subsequent mention of this item was deaccented, in line with empirical results (Brown 1983) suggesting that listeners associate accented items with NEW information and deaccented items with given information.

In sum, the original accent assignment algorithm in NewSpeak operated as follows:

- If the current word was a cue phrase, it received pitch accent assignment based upon its phrasal position, as described in (Litman and Hirschberg 1990);
- If the current word was a closed class item, it was deaccented;
- If the current word's root was in local or global focus, it was deaccented;
- If compound-stress assignment indicated that the word should be deaccented, it was deaccented;
- Otherwise, the word was accented.

While this algorithm proved fairly successful at producing natural-sounding speech from text such as AP News stories, it was clear from observations of human speech that not all that is 'given' is deaccented. It was also clear that NewSpeak's model of attentional state was insufficient to model even the given/new distinction. In particular, the domain over which items remain given – and the process by which they lose that 'givenness' — are open research questions.

To refine this original approach to accent assignment, to develop some evaluation procedures for prosodic assignment in speech synthesis, and to contribute more generally to research on models of discourse, NewSpeak's pitch accent assignment algorithm is being tested against prosodic labeling of natural speech. Certain interesting disparities have been found so far between NewSpeak output and labeled read speech.

Some of the deficiencies in the original algorithm appear due to an over-simplistic mapping between part-of-speech information and pitch accent. Some lexical items commonly counted as closed-class items regularly receive pitch accents. More difficult to correct are inadequacies in the discourse model. The simple identification of given information with deaccented lexical items and new information with accented items is clearly inadequate. Items which are quite plausibly deaccent*able* in terms of their information status are frequently not deaccent*ed*. And the behavior of nominal compounds with respect to both the imparting of 'givenness' to subsequent discourse and the consequences of 'givenness' on accent strategies for the compound, reveals considerable complexity. However, whether these problems can be addressed simply by refining the collection, representation, and updating of the attentional state, or must be supplemented by the representation of additional discourse features such as focus of attention and contrast remains to be determined.

Comparisons with Recorded Speech

The corpus being examined for prosodic comparison is the FM Radio Newscasting Database, a series of studio recordings of newscasts provided by National Public Radio Station WBUR in association with Boston University, which is being collected by SRI International (Patti Price), Boston University (Mari Ostendorf), and MIT (Stefanie Shattuck-Hufnagel). The prosodic analysis of this data which serves as the basis for the discussion below employs Pierrehumbert's (Pierrehumbert 1980) description of English intonation. In this system, intonational contours are described as sequences of low (L) and high (H) tones in the FUNDAMENTAL FREQUENCY CONTOUR (f0). A phrase's TUNE is represented as a sequence of one or more PITCH ACCENT(S), a PHRASE ACCENT, and a BOUNDARY TONE. For Pierrehumbert, there are six types of pitch accent in English, two simple tones – high and low – and four complex ones. The high tone, the most frequently used accent, comes out as a peak on the accented syllable and is represented as 'H*'; the 'H' indicates a high tone, and the '*' that the tone is aligned with a stressed syllable. Low (L*) accents occur much lower in the pitch

range than H* and are phonetically realized as local f0 minima. Complex accents have two tones, one of which is aligned with the stress. Using the diacritic '*' to indicate this alignment, these accents can be represented as L*+H, L+H*, H*+L, and H+L*. A well-formed INTERMEDIATE PHRASE consists of one or more pitch accents, plus a simple H or L tone which characterizes the phrase accent. The phrase accent spreads over the material between the last pitch accent of the current intermediate phrase and the beginning of the next — or the end of the utterance. INTONATIONAL PHRASES are composed of one of more such intermediate phrases plus a BOUNDARY TONE, which may also be H or L and is indicated by '%'. It falls exactly at the phrase boundary.

A sample of this prosodic labeling for the FM Radio database appears below:

```
Wanted.  Chief Justice of the  Massachusetts
 H*   LL% H*+L H*+L    -    -  L+H*
Supreme Court.   In April(1) the SJC's current
   H*+L  H*  LL% L* H*          -   -    H*
leader Edward Hennessy     reaches the mandatory
 -     H*      H*     LH%   H*        -  H*
retirement age of seventy  and a successor is
   H*      -   -  L+H*   L -   -    L+H*   -
expected to be named in March.   It may   be  the
   H*     - -  H*   -  H*+L LL% -  L+H* H*+L   -
most important appointment Governor Michael
 H*    L+H*    H*             -        H*
Dukakis makes  during the remainder of his
   H*   L+H* L  -       -   L+H*    -   -
administration   and     one of the toughest.   As
         H*   L L+H* L  L*  -   -   H+L*    LH% -
W BUR's Margo Millnicove reports, Hennessy(2) will
H* -     H*    H*           H*  L  H*          -
be a  hard  act to follow.
 - -  H*+L H*+L  -  H*   LL%
```

Results of Initial Comparisons

For experimental purposes, structural information used for accent assignment was limited to the part-of-speech tagging and NP identification available from Church's tagger (Church 1988). Also, NewSpeak's initial model of attentional state was replaced by one which allows procedures for constructing the attentional state to be varied systematically, and also permits additional sources of discourse information, such as FOCUS OF ATTENTION ('what attention is now to be focussed upon') to interact with given/new status and word class in pitch accent assignment. In the current implementation, such features can be added or subtracted from the model and methods of calculating their values can also be varied. The goal of such variation is of course the determination of which features, and which methods of calculating them, best predict pitch accent strategies in actual speech.

Initial comparisons have varied the following:

- how items are divided into closed/open-class by part-of-speech;
- which classes of items are considered in developing the attentional state model, e.g. all open-class items, nouns only, and so on;
- the mapping between word root and givenness (e.g., should mentioning *helpful* make *help* 'given'? *unhelpful*?);
- whether the attentional state should be structured as a simple stack of sets, or whether order should be imposed within the focal spaces, or whether distance between focal spaces on the stack should be important;
- how the attentional stack should be updated (e.g., by phrase, sentence, paragraph boundary or something else);
- whether nominal compounds function as ANAPHORIC ISLANDS for accenting purposes, as they do to some extent for pronominalization (i.e., whether the utterance of a compound nominal licenses the subsequent deaccenting of its subparts or not);
- whether notions of focus of attention and contrast, insofar as they can be inferred from unrestricted text, improve the model of accent assignment, and, if so, how they interact with word class and given/new distinctions.

Thus far, results from testing variations of NewSpeak's algorithm on two samples (each about five minutes long) from the FM Radio corpus suggest certain refinements of the original algorithm. Some of these are discussed below. However, it should be stressed that analysis of a much larger amount of labeled speech will be necessary to demonstrate their usefulness — and will also permit the analysis of interactions among the structural and discourse features described below.

Even from such slim data, it appears that the simple mapping between closed-class and deaccentuation employed in most text-to-speech systems must be modified. Word classes properly treated as closed for some purposes may nonetheless commonly be accented. In a more sophisticated variant of the NewSpeak's accent assignment algorithm, closed-class items are divided into three categories. Possessive pronouns (including *wh*-pronouns), definite and indefinite articles, copulas, coordinating and subordinating conjunctions, existential *there*, *have*, accusative pronouns and *wh*-adverbials, most prepositions, positive modals, positive *do*, as well as some particular adverbials like *ago*, nominative and accusative *it* and nominative *they*, and some nominal pronouns (e.g. *something*) are identified as 'closed, deaccented'. And certain of these classes are marked for further reduction in synthesis via CLITICIZATION, involving the removal of adjacent word boundaries and vowel reduction. 'Closed, accented' items, on the other hand, include the negative article, negative modals, negative *do*, most nominal

pronouns, most nominative and all reflexive pronouns, pre- and post-qualifiers (e.g. *quite*), pre-quantifiers (e.g. *all*), post-determiners (e.g. *next*), nominal adverbials (e.g. *here*), interjections, particles, most *wh*-words, plus some prepositions (e.g. *despite, unlike*). Other word classes (adjectives, adverbials,common and proper nouns, verbs) are deemed 'open'. For purposes of acquiring given/new information, only open-class items are considered, although how much of this category to consider is also subject to variation.

The collection and manipulation of the attentional state representation has been varied experimentally in the following ways: both global and local focus representations have been manipulated independently such that the global focus space may be set, and the local focus spaces updated, by the orthographic phrase, the sentence, or the paragraph. So, for example, the global space may be set after the first phrase, sentence or paragraph of a text. The local stack can be updated independently at the end of each phrase, sentence or paragraph – although cue phrases will push or pop the stack as well. For the current experiments, the best results to date have come when the global space is defined to be the first full sentence of the text and the local attentional stack is updated by paragraph. The content of both global and local focus spaces have also been varied systematically by word class, so that all open-class words, nouns only, or nouns plus some combination of verbs and modifiers are allowed to affect — and be affected by — the attentional state representation. Preliminary results, which again should be taken as suggestive only, indicate that focal spaces defined in terms of roots of all content words, rather than nominals only, or even all nonverb roots, provide the best accent prediction.

Finally, some experimentation has been done to relate the accenting of items currently in local focus with structural and discourse-based indicators of contrastiveness. For example, the referential strategy of PROPER-NAMING (Sanford *et al.*), in which the use of proper names was found to focus attention, may provide an explanation for an observed propensity in the FM Radio data for accenting proper names on subsequent mention (e.g. the pitch accent on *Hennessy* ((2) in the FM database sample). It is conjectured that such referential behavior might indicate the speaker's attempt to focus attention upon persons recently mentioned, when other focii have intervened since their introduction. This strategy, together with others which can be inferred from surface and syntactic features of the text, such as the preposing of adverbials and of prepositional phrases (e.g., *In April* ((1) in the sample), and the reintroduction of items in global but not in local focus, are hypothesized as predictors of contrastive focussing accents. Such considerations often appear to account for the accenting of items which seem clearly 'given' and thus potentially deaccentable. Hypothesizing such focussing behavior appears to be useful enough in accent prediction to warrant the investigation of other structural and discourse correlates of focus in future work.

Discussion

This paper has described the pitch accent assignment strategy employed in NewSpeak, an interface to the Bell Labs Text-to-Speech System, which employs a hierarchical representation of the attentional structure of the discourse, together with more traditional syntactic information, to assign intonational features in the synthesis of unrestricted text. It has also described experiments currently being performed to refine that algorithm, by modifying traditional uses of word class, key word, and surface position information, and by varying the construction of and interaction between the components of a model of attentional state. The testing of various discourse models against pitch accent placement in actual speech, should also add to our set of evaluation criteria for such models.

From a theoretical point of view, such analysis should bring us closer to understanding how to model pitch accent in human speech. However, real progress will depend upon the availability of large amounts of prosodically labeled data. From a more practical point of view, an immediate use of the accent assignment algorithm being developed will be to facilitate just such data analysis, providing hypotheses about prosodic features which can then be post-edited by hand — to speed up the labeling process. The resulting labeled speech can then of course be used for further training of the algorithm.

While the use of higher level discourse information to inform algorithms for pitch accent assignment appears to be a useful strategy for modeling accent assignment in natural speech, it may indeed turn out not to be desirable to emulate natural speech in synthetic speech. The work described above assumes, however, that whatever variation eventually emerges as desirable between synthetic speech and human speech should clearly be intentional and principled rather than chance. Proving that one speech synthesizer is preferable to another, or that one prosodic strategy is to be favored over another, in terms of simple human preference is notoriously difficult to accomplish. So, comparison of the output of algorithms used to assign intonational features in synthetic speech with prosodic features in natural speech would appear to be our most effective test of whether or not we are making actual improvements in the prosody of speech synthesis.

Acknowledgments

Thanks to Ken Church, Rick Omanson, and Richard Sproat for helpful comments.

References

Altenberg, B. 1987. *Prosodic Patterns in Spoken English: Studies in the Correlation between Prosody*

and Grammar for Text-to-Speech Conversion, volume 76 of *Lund Studies in English*. Lund University Press, Lund.

Bolinger, D. 1961. Contrastive accent and contrastive stress. *Language*, 37:83–96.

Bolinger, D. 1972. Accent is predictable (if you're a mindreader). *Language*, 48:633–644.

Brennan, S. E., Friedman, M. W., and Pollard, C. J. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting*, pages 155–162, Buffalo. Association for Computational Linguistics.

Brown, G. 1983. Prosodic structure and the given/new distinction. In Ladd, D. R. and Cutler, A., editors, *Prosody: Models and Measurements*. Springer Verlag, Berlin.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, C., editor, *Subject and Topic*. Academic Press, New York.

Church, K. W. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pages 136–143, Austin. Association for Computational Linguistics.

Clark, H. H. and Clark, E. V. 1977. *Psychology and Language*. Harcourt, Brace, Jovanovich, Inc.

Cruttenden, A. 1986. *Intonation*. Cambridge University Press, Cambridge UK.

Crystal, D. 1975. *The English Tone of Voice: Essays in Intonation, Prosody, and Paralanguage*. Edward Arnold, London.

Danlos, L., LaPorte, E., and Emerard, F. 1986. Synthesis of spoken message from semantic representation. In *Proceedings of COLING-86*, pages 599–604. COLING.

Davis, J. R. and Hirschberg, J. 1988. Assigning intonational features in synthesized spoken directions. In *Proceedings of the 26th Annual Meeting*, pages 187–193, Buffalo. Association for Computational Linguistics.

Grosz, B. and Sidner, C. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Grosz, B., Joshi, A., and Weinstein, S. June 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting*, pages 44–50, Cambridge MA. Association for Computational Linguistics.

Halliday, M. A. K. and Hassan, R. 1976. *Cohesion in English*. Longman.

Hindle, D. M. 1989. Acquiring disambiguation rules from text. In *Proceedings of the 27th Annual Meeting*, pages 118–125, Vancouver. Association for Computational Linguistics.

Jackendoff, R. S. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge MA.

Liberman, M. and Sproat, R. Forthcoming. The stress and structure of modified noun phrases in English. In Sag, I., editor, *Lexical Matters*. University of Chicago Press.

Litman, D. and Hirschberg, J. August 1990. Disambiguating cue phrases in text and speech. In *Proceedings of COLING90*, Helsinki. COLING.

Olive, J. P. and Liberman, M. Y. 1985. Text to speech – an overview. *Journal of the Acoustic Society of America, Suppl. 1*, 78(Fall):s6.

Pierrehumbert, J. B. 1980. *The Phonology and Phonetics of English Intonation*. PhD thesis, Massachusetts Institute of Technology.

Prince, E. 1981. Toward a taxonomy of given-new information. In Cole, P., editor, *Radical Pragmatics*. Academic Press, New York.

Quirk, R., Svartvik, J., Duckworth, A. P., Rusiecki, J. P. L., and Colin, A. J. T. 1964. Studies in the correspondence of prosodic to grammatical features in English. In *Proceedings of the Ninth International Congress*, pages 679–691. International Congress of Linguists.

Rooth, M. 1985. *Association with Focus*. PhD thesis, University of Massachusetts, Amherst MA.

Sanford, A. J., Garrod, S. C., Moar, K., and Al-Ahmar, H. 1985. Naming, role-descriptions, and main and secondary characters in discourse comprehension. Reported in Sanford, A. J. Aspects of pronoun interpretation: Evaluation of search formulations of inference. In Rickheit, G. and Strohner, H., editors, *Inferences in Text Processing*, pages 183–204. North-Holland, Amsterdam.

Sidner, C. L. 1983. Focusing in the comprehension of definite anaphora. In Brady, M., editor, *Computational Models of Discourse*. MIT Press, Cambridge MA.

Silverman, K. 1987. *The Structure and Processing of Fundamental Frequency Contours*. PhD thesis, Cambridge University, Cambridge UK.

Walker, M. A. 1989. Evaluating discourse processing algorithms. In *Proceedings of the 27th Annual Meeting*, pages 251–261, Vancouver. Association for Computational Linguistics.

Young, S. J. and Fallside, F. 1979. Speech synthesis from concept: A method for speech output from information systems. *Journal of the Acoustic Society of America*, 66(3):685–695.