Constituent-based Accent Prediction

Christine H. Nakatani
AT&T Labs – Research
180 Park Avenue, Florham Park NJ 07932-0971, USA
email: *chn@research.att.com*

Abstract

Near-perfect automatic accent assignment is attainable for citation-style speech, but better computational models are needed to predict accent in extended, spontaneous discourses. This paper presents an empirically motivated theory of the discourse focusing nature of accent in spontaneous speech. Hypotheses based on this theory lead to a new approach to accent prediction, in which patterns of deviation from citation form accentuation, defined at the constituent or noun phrase level, are automatically learned from an annotated corpus. Machine learning experiments on 1031 noun phrases from eighteen spontaneous direction-giving monologues show that accent assignment can be significantly improved by up to 4%-6% relative to a hypothetical baseline system that would produce only citation-form accentuation, giving error rate reductions of 11%-25%.

1 Introduction

In speech synthesis systems, near-perfect (98%) accent assignment is automatically attainable for read-aloud, citation-style speech (Hirschberg, 1993). But for unrestricted, extended spontaneous discourses, highly natural accentuation is often achieved only by costly human post-editing. A better understanding of the effects of discourse context on accentual variation is needed not only to fully model this fundamental prosodic feature for text-to-speech (TTS) synthesis systems, but also to further the integration of prosody into speech understanding and concept-to-speech (CTS) synthesis systems at the appropriate level of linguistic representation.

This paper presents an empirically motivated theory of the discourse focusing function of accent. The theory describes for the first time the interacting contributions to accent prediction made by factors related to the local and global attentional status of discourse referents in a discourse model (Grosz and Sidner, 1986). The ability of the focusing features to predict accent for a blind test corpus is examined using machine learning. Because attentional status is a property of referring expressions, a novel approach to accent prediction is proposed to allow for the integration of word-based and constituent-based linguistic features in the models to be learned.

The task of accent assignment is redefined as the prediction of patterns of deviation from citation form accentuation. Crucially, these deviations are captured at the constituent level. This task redefinition has two novel properties: (1) it bootstraps directly on knowledge about citation form or so-called "context-independent" prosody embodied in current TTS technology; and (2) the abstraction from word to constituent allows for the natural integration of focusing features into the prediction methods.

Results of the constituent-based accent prediction experiments show that for two speakers from a corpus of spontaneous direction-giving monologues, accent assignment can be improved by up to 4%-6% relative to a hypothetical baseline system that would produce only citation-form accentuation, giving error rate reductions of 11%-25%.

2 Accent and attention

Much theoretical work on intonational meaning has focused on the association of accent with NEW information, and lack of accent with GIVEN information, where given and new are defined with respect to whether or not the information is already represented in a discourse model. While this association reflects a general tendency (Brown, 1983), empirical studies on longer discourses have shown this simple dichotomy cannot explain important subclasses of expressions, such as accented pronouns, cf. (Terken, 1984; Hirschberg, 1993).

We propose a new theory of the relationship between accent and attention, based on an enriched taxonomy of given/new information status provided by both the LOCAL (centering) and GLOBAL (focus stack model) attentional state models in Grosz and Sidner's discourse modeling theory (1986).

Analysis of a 20-minute spontaneous story-telling monologue[1] identified separate but interacting contributions of grammatical function, form of referring expression and accentuation[2] in conveying the attentional status of a discourse referent. These interactions can be formally expressed in the framework of attentional modeling by the following principles of interpretation:

- The LEXICAL FORM OF A REFERRING EXPRESSION indicates the level of attentional processing, i.e., pronouns involve *local* focusing while full lexical forms involve *global* focusing (Grosz et al., 1995).
- The GRAMMATICAL FUNCTION of a referring expression reflects the local attentional status of the referent, i.e., subject position generally holds the highest ranking member of the forward-looking centers list (Cf list), while direct object holds the next highest ranking member of the Cf list (Grosz et al., 1995; Kameyama, 1985).
- The ACCENTING of a referring expression serves as an inference cue to shift attention to a new backward-looking center (Cb), or to mark the global (re)introduction of a referent; LACK OF ACCENT serves as an inference cue to maintain attentional focus on the Cb, Cf list members or global referents (Nakatani, 1997).

The third principle concerning accent interpretation defines for the first time how accent serves uniformly to *shift* attention and lack of accent serves to *maintain* attention, at *either* the local or global level of discourse structure. This principle describing the discourse focusing functions of accent directly explains 86.5% (173/200) of the referring expressions in the spontaneous narrative, as shown in Table 1. If performance factors (e.g. repairs, interruptions) and special discourse situations (e.g. direct quotations) are also considered accounted for, then coverage increases to 96.5% (193/200).

SUBJECT PRONOUNS (N=111)		
25	**prominent**	23%
	16	*shift in Cb*
	6	contrast
	3	emphasis
86	**nonprominent**	77%
	75	*continue or resume Cb*
	3	repair
	2	dialogue tag
	1	interruption from interviewer
	5	unaccounted for
DIRECT OBJECT PRONOUNS (N=15)		
1	**prominent**	7%
	1	contrast
14	**nonprominent**	93%
	10	*maintain non-Cb in Cf list*
	3	inter-sentential anaphora
	1	repair
SUBJECT EXPLICIT FORMS (N=54)		
49	**prominent**	91%
	44	*introduce new global ref as Cp*
	2	quoted context
	1	repair
	2	unaccounted for
5	**nonprominent**	9%
	2	top-level global focus
	1	quoted context
	1	repair
	1	interruption from interviewer
DIRECT OBJECT EXPLICIT FORMS (N=20)		
11	**prominent**	55%
	11	*introduce new global referent*
9	**nonprominent**	45%
	7	*maintain ref in global focus*
	2	quoted context

Table 1: Coverage of narrative data. The discourse focusing functions of accent appear in italics.

3 Constituent-based experiments

To test the generality of the proposed account of accent and attention, the ability of local and global focusing features to predict accent for a blind corpus is examined using machine learning. To rigorously assess the potential gains to be had from these attentional features, we consider them in combination with lexical and syntactic features identified in the literature as strong predictors of accentuation (Altenberg, 1987; Hirschberg, 1993; Ross et al., 1992). Previous studies, nonetheless, were aimed at predicting word accentuation, and so the features we borrow are being tested for the first time in learning the abstract accentuation patterns of syntactic constituents, specifically noun phrases (NPs).

3.1 Methods

Accent prediction models are learned from a corpus of unrestricted, spontaneous direction-giving monologues from the Boston Directions Corpus (Nakatani et al., 1995). Eighteen spontaneous direction-giving monologues are analyzed from two American English speakers, H1 (male) and H3 (female). The monologues range from 43 to 631 words in length, and comprise 1031 referring expressions made up of 2020 words. Minimal, non-recursive

[1]The narrative was collected by Virginia Merlini.

[2]Accented expressions are identified by the presence of PITCH ACCENT (Pierrehumbert, 1980).

Accent class	TTS-assigned accenting	Actual accenting
citation	*a* LITTLE SHOPPING AREA *we*	*a* LITTLE SHOPPING AREA *we*
supra	*one* *a* PRETTY *nice* AMBIANCE	ONE *a* PRETTY NICE AMBIANCE
reduced	*the* GREEN LINE SUBWAY YET ANOTHER RIGHT TURN	*the* GREEN *Line* SUBWAY *yet* ANOTHER RIGHT TURN
shift	*a* VERY FAST FIVE MINUTE *lunch*	*a* VERY FAST FIVE *minute* LUNCH

Table 3: Examples of citation-based accent classes. Accented words appear in boldface.

NP constituents, referred to as BASENPs, are automatically identified using Collins' (1996) lexical dependency parser. In the following complex NP, baseNPs appear in square brackets: *[the brownstone apartment building] on [the corner] of [Beacon and Mass Ave]*. BaseNPs are semi-automatically labeled for lexical, syntactic, local focus and global focus features. Table 2 provides summary corpus statistics. A rule-based machine learning program,

Corpus measure	H1	H3	Total
total no. of words	2359	1616	3975
baseNPs	621	410	1031
words in baseNPs	1203	817	2020
% words in baseNPs	51.0%	50.6%	50.8%

Table 2: Word and baseNP corpus measures.

Ripper (Cohen, 1995), is used to acquire accent classification systems from a training corpus of correctly classified examples, each defined by a vector of feature values, or predictors.[3]

3.2 Citation-based Accent Classification

The accentuation of baseNPs is coded according to the relationship of the actual accenting (i.e. accented versus unaccented) on the words in the baseNP to the accenting predicted by a TTS system that received each sentence in the corpus in isolation. The actual accenting is determined by prosodic labeling using the ToBI standard (Pitrelli et al., 1994). Word accent predictions are produced by the Bell Laboratories NewTTS system (Sproat, 1997). NewTTS incorporates complex nominal accenting rules (Sproat, 1994) as well as general, word-based accenting rules (Hirschberg, 1993). It is assumed for the purposes of this study that NewTTS generally assigns citation-style accentuation when passed sentences in isolation.

For each baseNP, one of the following four accenting patterns is assigned:

- CITATION FORM: exact match between actual and TTS-assigned word accenting.
- SUPRA: one or more accented words are predicted unaccented by TTS; otherwise, TTS predictions match actual accenting.
- REDUCED: one or more unaccented words are predicted accented by TTS; otherwise, TTS predictions match actual accenting.
- SHIFT: at least one accented word is predicted unaccented by TTS, and at least one unaccented word is predicted accented by TTS.

Examples from the Boston Directions Corpus for each accent class appear in Table 3.

Table 4 gives the breakdown of coded baseNPs by accent class. In contrast to read-aloud citation-style

Accent class	H1 baseNPs N	H1 baseNPs %	H3 baseNPs N	H3 baseNPs %
citation	471	75.8%	247	60.2%
supra	73	11.8%	68	16.6%
reduced	68	11.9%	83	20.2%
shift	9	1.4%	12	2.9%
total	621	100%	410	100%

Table 4: Accent class distribution for all baseNPs.

speech, in these unrestricted, spontaneous monologues, 30% of referring expressions do not bear citation form accentuation. The citation form accent percentages serve as the baseline for the accent prediction experiments; correct classification rates above 75.8% and 60.2% for H1 and H3 respectively would represent performance above and beyond the

[3]Ripper is similar to CART (Breiman et al., 1984), but it directly produces IF-THEN logic rules instead of decision trees and also utilizes incremental error reduction techniques in combination with novel rule optimization strategies.

state-of-the-art citation form accentuation models, gained by direct modeling of cases of supra, reduced or shifted constituent-based accentuation.

3.3 Predictors

3.3.1 Lexical features

The use of set features, which are handled by Ripper, extends lexical word features to the constituent level. Two set-valued features, BROAD CLASS SEQUENCE and LEMMA SEQUENCE, represent lexical information. These features consist of an ordered list of the broad class part-of-speech (POS) tags or word lemmas for the words making up the baseNP. For example, the lemma sequence for the NP, *the Harvard Square T stop*, is {the, Harvard, Square, T, stop}. The corresponding broad class sequence is {determiner, noun, noun, noun, noun}. Broad class tags are derived using Brill's (1995) part-of-speech tagger, and word lemma information is produced by NewTTS (Sproat, 1997).

POS information is used to assign accenting in nearly all speech synthesis systems. Initial word-based experiments on our corpus showed that broad class categories performed slightly better than both the function-content distinction and the POS tags themselves, giving 69%-81% correct word predictions (Nakatani, 1997).

3.3.2 Syntactic constituency features

The CLAUSE TYPE feature represents global syntactic constituency information, while the BASENP TYPE feature represents local or NP-internal syntactic constituency information. Four clause types are coded: *matrix, subordinate, predicate complement* and *relative*. Each baseNP is semi-automatically assigned the clause type of the lowest level clause or nearest dominating clausal node in the parse tree, which contains the baseNP. As for baseNP types, the baseNP type of baseNPs not dominated by any NP node is SIMPLE-BASENP. BaseNPs that occur in complex NPs (and are thus dominated by at least one NP node) are labeled according to whether the baseNP contains the head word for the dominating NP. Those that are dominated by only one NP node and contain the head word for the dominating NP are HEAD-BASENPS; all other NPs in a complex NP are CHILD-BASENPS. Conjoined noun phrases involve additional categories of baseNPs that are collapsed into the CONJUNCT-BASENP category. Table 5 gives the distributions of baseNP types.

Focus projection theories of accent, e.g. (Gussenhoven, 1984; Selkirk, 1984), would predict a large role for syntactic constituency information in determining accent, especially for noun phrase constituents. Empirical evidence for such a role, however, has been weak (Altenberg, 1987).

baseNP type	H1		H3	
	N	%	N	%
simple	447	72.0%	280	68.3%
head	61	9.8%	46	11.2%
child	74	11.9%	65	15.9%
conjunct	39	6.3%	19	4.5%
total	621	100%	410	100%

Table 5: Distribution of baseNP types for all baseNPs.

3.3.3 Local focusing features

The local attentional status of baseNPs is represented by two features commonly used in centering theory to compute the Cb and the Cf list, GRAMMATICAL FUNCTION and FORM OF EXPRESSION (Grosz et al., 1995). Hand-labeled grammatical functions include *subject, direct object, indirect object, predicate complement, adjunct*. Form of expression feature values are *adverbial noun, cardinal, definite NP, demonstrative NP, indefinite NP, pronoun, proper name, quantifier NP, verbal noun*, etc.

3.3.4 Global focus feature

The global focusing status of baseNPs is computed using two sets of analyses: discourse segmentations and coreference coding. Expert discourse structure analyses are used to derive CONSENSUS SEGMENTATIONS, consisting of discourse boundaries whose coding all three labelers agreed upon (Hirschberg and Nakatani, 1996). The consensus labels for segment-initial boundaries provide a linear segmentation of a discourse into discourse segments. Coreferential relations are coded by two labelers using DTT (Discourse Tagging Tool) (Aone and Bennett, 1995). To compute coreference chains, only the relation of strict coference is used. Two NPs, np1 and np2, are in a strict coreference relationship, when np2 occurs after np1 in the discourse and realizes the same discourse entity that is realized by np1. Reference chains are then automatically computed by linking noun phrases in strict coference relations into the longest possible chains.

Given a consensus linear segmentation and reference chains, global focusing status is determined. For each baseNP, if it does not occur in a reference chain, and thus is realized only once in the dis-

course, it is assigned the SINGLE-MENTION focusing status. The remaining statuses apply to baseNPs that do occur in reference chains. If a baseNP in a chain is not previously mentioned in the discourse, it is assigned the FIRST-MENTION status. If its most recent coreferring expression occurs in the current segment, the baseNP is in IMMEDIATE focus; if it occurs in the immediately previous segment, the baseNP is in NEIGHBORING focus; if it occurs in the discourse but not in either the current or immediately previous segments, then the baseNP is assigned STACK focus.

4 Results

4.1 Individual features

Experimental results on individual features are reported in Table 4.1 in terms of the average percent correct classification and standard deviation.[4]

A trend emerges that lexical features (i.e. word lemma and broad class sequences, and form of expression) enable the largest improvements in classification, e.g. 2.7% and 2.3% for H1 using broad class sequence and form of expression information respectively. These results suggest that the abstract level of lexical description supplied by form of expression does the equivalent work of the lower-level lexical features. Thus, for CTS, accentuation class might be predicted when the more abstract form of expression information is known, and need not be delayed until the tactical generation of the expression is completed. Conversely, for TTS, simple corpus analysis of lemma and POS sequences may perform as well as higher-level lexical analysis.

Experiment	H1	H3
Lexical		
Broad cl seq	78.58 ± 1.30	59.51 ± 2.72
Lemma seq	*80.05 ± 1.85*	62.93 ± 2.68
Syntactic		
baseNP type	75.86 ± 2.52	60.24 ± 2.97
Clause type	75.85 ± 1.14	60.24 ± 3.49
Local focus		
Gram fn	75.83 ± 1.93	62.68 ± 2.74
Form of expr	78.10 ± 1.54	61.95 ± 1.89
Global focus		
Global focus	75.85 ± 2.07	—
Baseline	75.8	60.2

Table 6: Average percentages correct classification and standard deviations for individual feature experiments.

4.2 Combinations of classes of features

Experiments on combinations of feature classes are reported in Table 7. The average classification rate of 63.17% for H3 on the local focus and lexical feature class model, is the best obtained for all H3 experiments, increasing prediction accuracy by nearly 3%. The highest classification rate for H1 is 79.22% for the model including local and global focus, and lexical and syntactic feature classes, showing an improvement of 3.4%. These results, however, do not attain significance.

Experiment	H1	H3
Local/syntax	77.61 ± 1.39	60.98 ± 2.60
Local/lex	78.74 ± 1.48	63.17 ± 1.90
Local/lex/syntax	79.06 ± 1.53	61.95 ± 2.27
Local/global	78.11 ± 1.28	—
Loc/glob/lex/syn	79.22 ± 1.96	—
Baseline	75.8	60.2

Table 7: Average percentages correct classification and standard deviations for combination experiments.

4.3 Experiments on simple-baseNPs

Three sets of experiments that showed strong performance gains are reported for the non-recursive simple-baseNPs. These are: (1) word lemma sequence alone, (2) lemma and broad class sequences together, and (3) local focus and lexical features combined. Table 8 shows the accent class distribution for simple-baseNPs.

Accent class	H1 simple-baseNPs N	%	H3 simple-baseNPs N	%
citation	334	74.7	167	59.6
supra	62	13.9	47	16.8
reduced	46	10.3	56	0.20
shift	5	1.1	10	3.6
total	447	100	280	100

Table 8: Accent class distribution for simple-baseNPs.

Results appear in Table 9. For H3, the lemma sequence model delivers the best performance, 65.71%, for a 4.3% improvement over the baseline. The best classification rate of 80.93% for H1 on the local focus and lexical feature model represents a 6.23% gain over the baseline. These figures represent an 11% reduction in error rate for H3, and a

[4]Ripper experiments are conducted with 10-fold cross-validation. Statistically significant differences in the performance of two systems are determined by using the Student's curve approximation to compute confidence intervals, following Litman (1996). Significant results at $p < .05$ or stronger appear in italics.

25% reduction in error rate for H1, and are statistically significant improvements over the baseline.

Experiment	H1	H3
Lemma seq	*80.74 ± 1.87*	*65.71 ± 2.70*
Lemma, broad cl	*80.80 ± 1.41*	62.14 ± 2.58
Local/lexical	*80.93 ± 1.35*	63.21 ± 1.78
Baseline	74.7	59.6

Table 9: Average percentages correct classification and standard deviations for simple-baseNP experiments.

In the rule sets learned by Ripper for the H1 local focus/lexical model, interactions of the different features in specific rules can be observed. Two rule sets that performed with error rates of 13.6% and 13.7% on different cross-validation runs are presented in Figure 1.[5] Inspection of the rule sets

H1 local focus/lexical model rule set 1

reduced :- form of expr=proper name, broad class seq ~ det, lemma seq ~ Harvard.
supra :- broad class seq ~ adverbial.
supra :- gram fn=adjunct, lemma seq ~ this.
supra :- gram fn=adjunct, lemma seq ~ Cowperwaithe.
supra :- lemma seq ~ I.
default citation.

H1 local focus/lexical model rule set 2

reduced:- broad class seq ~ n, lemma seq ~ the, lemma seq ~ Square.
supra :- form of expr=adverbial.
supra :- gram fn=adjunct, lemma seq ~ Cowperwaithe.
supra :- lemma seq ~ this.
supra :- lemma seq ~ I.
default citation.

Figure 1: Highest performing learned rule sets for H1, local focus/lexical model.

reveals that there are few non-lexical rules learned. The exception seems to be the rule that adverbial noun phrases belong to the supra accent class. However, new interactions of local focusing features (grammatical function and form of expression) with lexical information are discovered by Ripper. It also appears that as suggested by earlier experiments, lexical features trade-off for one other as well as with form of expression information. In comparing the first rules in each set, for example, the clauses broad class seq ~ det and lemma seq ~ the substitute for one another. However, in the first rule set the less specific broad class constraint must be combined with another abstract constraint, form of expr=proper name, to achieve a similar description of a rule for reduced accentuation on common place names, such as *the Harvard Square T stop.*

[5] In the rules themselves, written in Prolog-style notation, the tilde character is a two-place operator, X ~ Y, signifying that Y is a member of the set-value for feature X.

5 Conclusion

Accent prediction experiments on noun phrase constituents demonstrated that deviations from citation form accentuation (supra, reduced and shift classes) can be directly modeled. Machine learning experiments using not only lexical and syntactic features, but also discourse focusing features identified by a new theory of accent interpretation in discourse, showed that accent assignment can be improved by up to 4%-6% relative to a hypothetical baseline system that would produce only citation-form accentuation, giving error rate reductions of 11%-25%. In general, constituent-based accentuation is most accurately learned from lexical information readily available in TTS systems. For CTS systems, comparable performance may be achieved using only higher level attentional features. There are several other lessons to be learned, concerning individual speaker, domain dependent and domain independent effects on accent modeling.

First, it is perhaps counterintuitively harder to predict deviations from citation form accentuation for speakers who exhibit a great deal of non-citation-style accenting behavior, such as speaker H3. Accent prediction results for H1 exceeded those for H3, although about 15% more of H3's tokens exhibited non-citation form accentuation. Finding the appropriate parameters by which to describe the prosody of individual speakers is an important goal that can be advanced by using machine learning techniques to explore large spaces of hypotheses.

Second, it is evident from the strong performance of the word lemma sequence models that deviations from citation-form accentuation may often be expressed by lexicalized rules of some sort. Lexicalized rules in fact have proven useful in other areas of natural language statistical modeling, such as POS tagging (Brill, 1995) and parsing (Collins, 1996). The specific lexicalized rules learned for many of the models would not have followed from any theoretical or empirical proposals in the literature. It may be that domain dependent training using au-

tomatic learning is the appropriate way to develop practical models of accenting patterns on different corpora. And especially for different speakers in the same domain, automatic learning methods seem to be the only efficient way to capture perhaps idiolectical variation in accenting.

Finally, it should be noted that notwithstanding individual speaker and domain dependent effects, domain independent factors identified by the new theory of accent and attention do contribute to experimental performance. The two local focusing features, grammatical function and form of referring expression, enable improvements above the citation-form baseline, especially in combination with lexical information. Global focusing information is of limited use by itself, but as may have been hypothesized, contributes to accent prediction in combination with local focus, lexical and syntactic features.

Acknowledgments

This research was supported by a NSF Graduate Research Fellowship and NSF Grants Nos. IRI-90-09018, IRI-93-08173 and CDA-94-01024 at Harvard University. The author is grateful to Barbara Grosz, Julia Hirschberg and Stuart Shieber for valuable discussion on this research; to Chinatsu Aone, Scott Bennett, Eric Brill, William Cohen, Michael Collins, Giovanni Flammia, Diane Litman, Becky Passonneau, Richard Sproat and Gregory Ward for sharing and discussing methods and tools; and to Diane Litman, Marilyn Walker and Steve Whittaker for suggestions for improving this paper.

References

B. Altenberg. 1987. *Prosodic Patterns in Spoken English: Studies in the Correlation Between Prosody and Grammar for Text-to-Speech Conversion.* Lund University Press, Lund, Sweden.

C. Aone and S. W. Bennett. 1995. Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd Annual Meeting*, Boston. Association for Computational Linguistics.

Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees.* Wadsworth and Brooks, Pacific Grove CA.

Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: a case study in part of speech tagging. *Computational Lingusitics.*

G. Brown. 1983. Prosodic structure and the Given/New distinction. In A. Cutler and D. R. Ladd, editors, *Prosody: Models and Measurements*, pages 67–78. Springer-Verlag, Berlin.

William A. Cohen. 1995. Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning.*

Michael John Collins. 1996. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics.*

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

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2), June.

Carlos Gussenhoven. 1984. *On the Grammar and Semantics of Sentence Accents.* Foris Publications, Dordrecht.

Julia Hirschberg and Christine Nakatani. 1996. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the 34th Annual Meeting of the ACL*, Santa Cruz. Association for Computational Linguistics.

Julia Hirschberg. 1993. Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence*, 63(1-2):305–340.

M. Kameyama. 1985. *Zero anaphora: the case in Japanese.* Ph.D. thesis, Stanford University.

Diane J. Litman. 1996. Cue phrase classification using machine learning. *Journal of Artificial Intelligence*, pages 53–94.

Christine H. Nakatani, Barbara Grosz, and Julia Hirschberg. 1995. Discourse structure in spoken language: studies on speech corpora. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Palo Alto, CA, March. American Association for Artificial Intelligence.

Christine H. Nakatani. 1997. *The Computational Processing of Intonational Prominence: a Functional Prosody Perspective.* Ph.D. thesis, Harvard University, Cambridge, MA, May.

Janet Pierrehumbert. 1980. *The Phonology and Phonetics of English Intonation.* Ph.D. thesis, Massachusetts Institute of Technology, September. Distributed by the Indiana University Linguistics Club.

John Pitrelli, Mary Beckman, and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, volume 2, pages 123–126, Yokohama, Japan.

K. Ross, M. Ostendorf, and S. Shattuck-Hufnagel. 1992. Factors affecting pitch accent placement. In *Proceedings of the 2nd International Conference on Spoken Language Processing*, pages 365–368, Banff, Canada, October.

E. Selkirk. 1984. *Phonology and Syntax.* MIT Press, Cambridge MA.

Richard Sproat. 1994. English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language*, 8:79–94.

Richard Sproat, editor. 1997. *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach.* Kluwer Academic, Boston.

J. Terken. 1984. The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, 27:269–289.