

# Modeling Local Context for Pitch Accent Prediction

**Shimei Pan**

Department of Computer Science  
Columbia University  
New York, NY, 10027, USA  
pan@cs.columbia.edu

**Julia Hirschberg**

AT&T Labs-Research  
Florham Park, NJ, 07932-0971, USA  
julia@research.att.com

## Abstract

Pitch accent placement is a major topic in intonational phonology research and its application to speech synthesis. What factors influence whether or not a word is made intonationally prominent or not is an open question. In this paper, we investigate how one aspect of a word's local context — its collocation with neighboring words — influences whether it is accented or not. Results of experiments on two transcribed speech corpora in a medical domain show that such collocation information is a useful predictor of pitch accent placement.

## 1 Introduction

In English, speakers make some words more intonationally prominent than others. These words are said to be *accented* or to bear *pitch accents*. Accented words are typically louder and longer than their unaccented counterparts, and their stressable syllable is usually aligned with an excursion in the *fundamental frequency*. This excursion will differ in shape according to the type of pitch accent. Pitch accent type, in turn, influences listeners' interpretation of the accented word or its larger syntactic constituent. Previous research has associated pitch accent with variation in various types of information status, including the *given/new* distinction, *focus*, and *contrastiveness*, inter alia. Assigning pitch accent in speech generation systems which employ speech synthesizers for output is thus critical to system performance: not only must one

convey meaning naturally, as humans would, but one must avoid conveying *mis*-information which reliance on the synthesizers' defaults may result in.

The speech generation work discussed here is part of a larger effort in developing an intelligent multimedia presentation generation system called MAGIC (Medical Abstract Generation for Intensive Care) (Dalal et al., 1996). In MAGIC, given a patient's medical record stored at Columbia Presbyterian Medical Center (CPMC)'s on-line database system, the system automatically generates a post-operative status report for a patient who has just undergone bypass surgery. There are two media-specific generators in MAGIC: a graphics generator which automatically produces graphical presentations from database entities, and a spoken language generator which automatically produces coherent spoken language presentations from these entities. The graphical and the speech generators communicate with each other on the fly to ensure that the final multimedia output is synchronized.

In order to produce natural and coherent speech output, MAGIC's spoken language generator models a collection of speech features, such as accenting and intonational phrasing, which are critical to the naturalness and intelligibility of output speech. In order to assign these features accurately, the system needs to identify useful correlates of accent and phrase boundary location to use as predictors. This work represents part of our efforts in identifying useful predictors for pitch accent placement.

Pitch accent placement has long been a research focus for scientists working on phonology, speech analysis and synthesis (Bolinger,

1989; Ladd, 1996). In general, syntactic features are the most widely used features in pitch accent predication. For example, part-of-speech is traditionally the most useful single pitch accent predictor (Hirschberg, 1993). Function words, such as prepositions and articles, are less likely to be accented, while content words, such as nouns and adjectives, are more likely to be accented. Other linguistic features, such as inferred given/new status (Hirschberg, 1993; Brown, 1983), contrastiveness (Bolinger, 1961), and discourse structure (Nakatani, 1998), have also been examined to explain accent assignment in large speech corpora. In a previous study (Pan and McKeown, 1998; Pan and McKeown, 1999), we investigated how features such as deep syntactic/semantic structure and word informativeness correlate with accent placement. In this paper, we focus on how local context influences accent patterns. More specifically, we investigate how word collocation influences whether nouns are accented or not.

Determining which nouns are accented and which are not is challenging, since part-of-speech information cannot help here. So, other accent predictors must be found. There are some advantages in looking only at one word class. We eliminate the interaction between part-of-speech and collocation, so that the influence of collocation is easier to identify. It also seems likely that collocation may have a greater impact on content words, like nouns, than on function words, like prepositions.

Previous researchers have speculated that word collocation affects stress assignment of noun phrases in English. For example, James Marchand (1993) notes how *familiar collocations change their stress, witness the American pronunciation of 'Little House' [in the television series Little House on the Prairie], where stress used to be on HOUSE, but now, since the series is so familiar, is placed on the LITTLE*. That is, for collocated words, stress shifts to the left element of the compound. However, there are numerous counter-examples: consider *apple PIE*, which retains a right stress pattern, despite the collocation. So, the extent to which collocational status affects accent patterns is still unclear.

Despite some preliminary investigation

(Lieberman and Sproat, 1992), word collocation information has not, to our knowledge, been successfully used to model pitch accent assignment; nor has it been incorporated into any existing speech synthesis systems. In this paper, we empirically verify the usefulness of word collocation for accent prediction. In Section 2, we describe our annotated speech corpora. In Section 3, we present a description of the collocation measures we investigated. Section 4 to 7 describe our analyses and machine learning experiments in which we attempt to predict accent location. In Section 8 we sum up our results and discuss plans for further research.

## 2 Speech Corpora

From the medical domain described in Section 1, we collected two speech corpora and one text corpus for pitch accent modeling. The speech corpora consist of one multi-speaker spontaneous corpus, containing twenty segments and totaling fifty minutes, and one read corpus of five segments, read by a single speaker and totaling eleven minutes of speech. The text corpus consists of 3.5 million words from 7375 discharge summaries of patients who had undergone surgery. The speech corpora only cover cardiac patients, while the text corpus covers a larger group of patients and the majority of them have also undergone cardiac surgery.

The speech corpora were first transcribed orthographically and then intonationally, using the ToBI convention for prosodic labeling of standard American English (Silverman et al., 1992). For this study, we used only binary accented/deaccented decisions derived from the ToBI tonal tier, in which location and type of pitch accent is marked. After ToBI labeling, each word in the corpora was tagged with part-of-speech, from a nine-element set: noun, verb, adjective, adverb, article, conjunction, pronoun, cardinal, and preposition. The spontaneous corpus was tagged by hand and the read tagged automatically. As noted above, we focus here on predicting whether nouns are accented or not.

## 3 Collocation Measures

We used three measures of word collocation to examine the relationship between collocation

and accent placement: WORD BIGRAM PREDICTABILITY, MUTUAL INFORMATION, and the DICE COEFFICIENT. While word predictability is not typically used to measure collocation, there is some correlation between word collocation and predictability. For example, if two words are collocated, then it will be easy to predict the second word from the first. Similarly, if one word is highly predictable given another word, then there is a higher possibility that these two words are collocated. Mutual information (Fano, 1961) and the Dice coefficient (Dice, 1945) are two standard measures of collocation. In general, mutual information measures uncertainty reduction or departure from independence. The Dice coefficient is a collocation measure widely used in information retrieval. In the following, we will give a more detailed definitions of each.

Statistically, bigram word predictability is defined as the log conditional probability of word  $w_i$ , given the previous word  $w_{i-1}$ :

$$Pred(w_i) = \log(Prob(w_i|w_{i-1}))$$

Bigram predictability directly measures the likelihood of seeing one word, given the occurrence of the previous word. Bigram predictability has two forms: absolute and relative. Absolute predictability is the value directly computed from the formula. For example, given four adjacent words  $w_{i-1}, w_i, w_{i+1}$  and  $w_{i+2}$ , if we assume  $Prob(w_i|w_{i-1}) = 0.0001$ ,  $Prob(w_{i+1}|w_i) = 0.001$ , and  $Prob(w_{i+2}|w_{i+1}) = 0.01$ , the absolute bigram predictability will be -4, -3 and -2 for  $w_i, w_{i+1}$  and  $w_{i+2}$ . The relative predictability is defined as the rank of absolute predictability among words in a constituent. In the same example, the relative predictability will be 1, 2 and 3 for  $w_i, w_{i+1}$  and  $w_{i+2}$ , where 1 is associated with the word with the lowest absolute predictability. In general, the higher the rank, the higher the absolute predictability. Except in Section 7, all the predictability measures mentioned in this paper use the absolute form.

We used our text corpus to compute bigram word predictability for our domain. When calculating the word bigram predictability, we first filtered uncommon words (words occurring 5 times or fewer in the corpus) then used

the Good-Turing discount strategy to smooth the bigram. Finally we calculated the log conditional probability of each word as the measure of its bigram predictability.

Two measures of mutual information were used for word collocation: POINTWISE MUTUAL INFORMATION, which is defined as :

$$I_1(w_{i-1}; w_i) = \log \frac{P_r(w_{i-1}, w_i)}{P_r(w_{i-1})P_r(w_i)}$$

and AVERAGE MUTUAL INFORMATION, which is defined as:

$$\begin{aligned} I_2(w_{i-1}; w_i) = & P_r(w_{i-1}, w_i) \log \frac{P_r(w_{i-1}, w_i)}{P_r(w_{i-1})P_r(w_i)} \\ & + P_r(w_{i-1}, \bar{w}_i) \log \frac{P_r(w_{i-1}, \bar{w}_i)}{P_r(w_{i-1})P_r(\bar{w}_i)} \\ & + P_r(\bar{w}_{i-1}, w_i) \log \frac{P_r(\bar{w}_{i-1}, w_i)}{P_r(\bar{w}_{i-1})P_r(w_i)} \\ & + P_r(\bar{w}_{i-1}, \bar{w}_i) \log \frac{P_r(\bar{w}_{i-1}, \bar{w}_i)}{P_r(\bar{w}_{i-1})P_r(\bar{w}_i)} \end{aligned}$$

The same text corpus was used to compute both mutual information measures. Only word pairs with bigram frequency greater than five were retained.

The Dice coefficient is defined as:

$$Dice(w_{i-1}, w_i) = \frac{2 \times P_r(w_{i-1}, w_i)}{P_r(w_{i-1}) + P_r(w_i)}$$

Here, we also use a cut off threshold of five to filter uncommon bigrams.

Although all these measures are correlated, one measure can score word pairs quite differently from another. Table 1 shows the top ten collocations for each metric.

In the predictability top ten list, we have pairs like *scarlet fever* where *fever* is very predictable from *scarlet* (in our corpus, *scarlet* is always followed by *fever*), thus, it ranks highest in the predictability list. Since *scarlet* can be difficult to predict from *fever*, these types of pairs will not receive a very high score using mutual information (in the top 5% in  $I_1$  sorted list and in the top 20% in  $I_2$  list) and Dice coefficient (top 22%). From this table, it is also quite clear that  $I_1$  tends to rank uncommon words high. All the words in the top ten  $I_1$  list have a frequency less than or equal

Pred	$I_1$	$I_2$	Dice
chief complaint	polymyalgia rheumatica	The patient	greenfield filter
cerebrospinal fluid	hemiside stepper	present illness	Guillain Barre
folic acid	Pepto Bismol	hospital course	Viet Nam
periprocedural complications	Glen Cove	p o	Neo Synephrine
normoactive bowel	hydrogen peroxide	physical exam	polymyalgia rheumatica
uric acid	Viet Nam	i d	hemiside stepper
postpericardiotomy syndrome	Neo Synephrine	coronary artery	Pepto Bismol
Staten Island	otitis media	postoperative day	Glen Cove
scarlet fever	Lo Gerfo	saphenous vein	present illness
pericardiotomy syndrome	Chlor Trimeton	medical history	chief complaint

Table 1: Top Ten Most Collocated Words for Each Measure

to seven (we filter all the pairs occurring fewer than six times).

Of the different metrics, only bigram predictability is a unidirectional measure. It captures how the appearance of one word affects the appearance of the following word. In contrast, the other measures are all bidirectional measures, making no distinction between the relative position of elements of a pair of collocated items. Among the bidirectional measures, point-wise mutual information is sensitive to marginal probabilities  $P_r(word_{i-1})$  and  $P_r(word_i)$ . It tends to give higher values as these probabilities decrease, independently of the distribution of their co-occurrence. The Dice coefficient, however, is not sensitive to marginal probability. It computes conditional probabilities which are equally weighted in both directions.

Average mutual information measures the reduction in the uncertainty, of one word, given another, and is totally symmetric. Since  $I_2(word_{i-1}; word_i) = I_2(word_i; word_{i-1})$ , the uncertainty reduction of the first word, given the second word, is equal to the uncertainty reduction of the second word, given the first word. Further more, because  $I_2(\overline{word_i}; \overline{word_{i-1}}) = I_2(word_i; word_{i-1})$ , the uncertainty reduction of one word, given another, is also equal to the uncertainty reduction of failing to see one word, having failed to see the other.

Since there is considerable evidence that prior discourse context, such as previous mention of a word, affects pitch accent decisions, it is possible that symmetric measures, such as mutual information and the Dice coefficient, may not model accent placement as well as asymmetric measures, such as bigram

predictability. Also, the bias of point-wise mutual information toward uncommon words can affect its ability to model accent assignment, since, in general, uncommon words are more likely to be accented (Pan and McKeown, 1999). Since this metric disproportionately raises the mutual information for uncommon words, making them more predictable than their appearance in the corpus warrants, it may predict that uncommon words are more likely to be *deaccented* than they really are.

#### 4 Statistical Analyses

In order to determine whether word collocation is useful for pitch accent prediction, we first employed Spearman’s rank correlation test (Conover, 1980).

In this experiment, we employed a unigram predictability-based baseline model. The unigram predictability of a word is defined as the log probability of a word in the text corpus. The maximum likelihood estimation of this measure is:

$$\log \frac{Freq(w_i)}{\sum_i Freq(w_i)}$$

The reason for choosing this as the baseline model is not only because it is context independent, but also because it is effective. In a previous study (Pan and McKeown, 1999), we showed that when this feature is used, it is as powerful a predictor as part-of-speech. When jointly used with part-of-speech information, the combined model can perform significantly better than each individual model. When tested on a similar medical corpus, this combined model also outperforms a comprehensive pitch accent model employed by the

Bell Labs' TTS system (Sproat et al., 1992; Hirschberg, 1993; Sproat, 1998), where discourse information, such as given/new, syntactic information, such as POS, and surface information, such as word distance, are incorporated. Since unigram predictability is context independent. By comparing other predictors to this baseline model, we can demonstrate the impact of context, measured by word collocation, on pitch accent assignment.

Table 2 shows that for our read speech corpus, unigram predictability, bigram predictability and mutual information are all significantly correlated ( $p < 0.001$ ) with pitch accent decision.<sup>1</sup> However, the Dice coefficient shows only a trend toward correlation ( $p < 0.07$ ). In addition, both bigram predictability and (pointwise) mutual information show a slightly stronger correlation with pitch accent than the baseline. When we conducted a similar test on the spontaneous corpus, we found that all but the baseline model are significantly correlated with pitch accent placement. Since all three models incorporate a context word while the baseline model does not, these results suggest the usefulness of context in accent prediction. Overall, for all the different measures of collocation, bigram predictability explains the largest amount of variation in accent status for both corpora. We conducted a similar test using trigram predictability, where two context words, instead of one, were used to predict the current word. The results are slightly worse than bigram predictability (for the read corpus  $r = -0.167$ ,  $p < 0.0001$ ; for the spontaneous  $r = -0.355$ ,  $p < 0.0001$ ). The failure of the trigram model to improve over the bigram model may be due to sparse data. Thus, in the following analysis, we focus on bigram predictability. In order to further verify the effectiveness of word predictability in accent prediction, we will show some examples in our speech corpora first. Then we will describe how machine learning helps to derive pitch accent prediction models using this feature. Finally, we show that both absolute predictability and relative predictability are useful for pitch accent prediction.

<sup>1</sup>Since pointwise mutual information performed consistently better than average mutual information in our experiment, we present results only for the former.

## 5 Word Predictability and Accent

In general, nouns, especially head nouns, are very likely to be accented. However, certain nouns consistently do not get accented. For example, Table 3 shows some collocations containing the word *cell* in our speech corpus. For each context, we list the collocated pair, its most frequent accent pattern in our corpus (upper case indicates that the word was accented and lower case indicates that it was deaccented), its bigram predictability (the larger the number is, the more predictable the word is), and the frequency of this accent pattern, as well as the total occurrence of the bigram in the corpus. In the first ex-

Word Pair	Pred(cell)	Freq
[of] CELL	-3.11	7/7
[RED] CELL	-1.119	2/2
[PACKED] cell	-0.5759	4/6
[BLOOD] cell	-0.067	2/2

Table 3: *cell* Collocations

ample, *cell* in [of] CELL is very unpredictable from the occurrence of *of* and always receives a pitch accent. In [RED] CELL, [PACKED] cell, and [BLOOD] cell, *cell* has the same semantic meaning, but different accent patterns: *cell* in [PACKED] cell and [BLOOD] cell is more predictable and deaccented, while in [RED] CELL it is less predictable and is accented. These examples show the influence of context and its usefulness for bigram predictability. Other predictable nouns, such as *saver* in CELL *saver* usually are not accented even when they function as head nouns. *Saver* is deaccented in ten of the eleven instances in our speech corpus. Its bigram score is -1.5517, which is much higher than that of CELL (-4.6394-3.1083 depending upon context). Without collocation information, a typical accent prediction system is likely to accent *saver*, which would be inappropriate in this domain.

## 6 Accent Prediction Models

Both the correlation test results and direct observations provide some evidence on the usefulness of word predictability. But we still need to demonstrate that we can successfully use this feature in automatic accent prediction. In order to achieve this, we used machine learning

Corpus	Read		Spontaneous	
	$r$	p-value	$r$	p-value
Baseline (Unigram)	$r = -0.166$	$p = 0.0002$	$r = -0.02$	$p = 0.39$
Bigram Predictability	$r = -0.236$	$p < 0.0001$	$r = -0.36$	$p < 0.0001$
Pointwise Mutual Information	$r = -0.185$	$p < 0.0001$	$r = -0.177$	$p < 0.0001$
Dice Coefficient	$r = -0.079$	$p = 0.066$	$r = -0.094$	$p < 0.0001$

Table 2: Correlation of Different Collocation Measures with Accent Decision

techniques to automatically build accent prediction models using bigram word predictability scores.

We used RIPPER (Cohen, 1995b) to explore the relations between predictability and accent placement. RIPPER is a classification-based rule induction system. From annotated examples, it derives a set of ordered if-then rules, describing how input features can be used to predict an output feature. In order to avoid overfitting, we use 5-fold cross validation. The training data include all the nouns in the speech corpora. The independent variables used to predict accent status are the unigram and bigram predictability measures, and the dependent variable is pitch accent status. We used a majority-based predictability model as our baseline (i.e. predict *accented*).

In the combined model, both unigram and bigram predictability are used together for accent prediction. From the results in Table 4, we see that the bigram model consistently outperforms the unigram model, and the combined model achieves the best performance.

To evaluate the significance of the improvements achieved by incorporating a context word, we use the standard error produced by RIPPER. Two results are statistically significant when the results plus or minus twice the standard error do not overlap (Cohen, 1995a). As shown in Table 4, for the read corpus, except for the unigram model, all the models with bigram predictability performed significantly better than the baseline model. However, the bigram model and the combined model failed to improve significantly over the unigram model. This may result from too small a corpus. For the spontaneous corpus, the unigram, bigram and the combined model all achieved significant improvement over the baseline. The bigram also performed significantly better than the unigram model. The combined model had the best performance. It

also achieved significant improvement over the unigram model.

The improvement of the combined model over both unigram and bigram models may be due to the fact that some accent patterns that are not captured by one are indeed captured by the other. For example, accent patterns for street names have been extensively discussed in the literature (Ladd, 1996). For example, *street* in phrases like (e.g. *FIFTH street*) is typically deaccented while *avenue* (e.g. *Fifth AVENUE*) is accented. While it seems likely that the conditional probability of  $P_r(\text{Street}|\text{Fifth})$  is no higher than that of  $P_r(\text{Avenue}|\text{Fifth})$ , the unigram probability of  $P_r(\text{street})$  is probably higher than that of  $P_r(\text{avenue})$ .<sup>2</sup> So, incorporating both predictability measures may tease apart these and similar cases.

## 7 Relative Predictability

In our previous analysis, we showed the effectiveness of absolute word predictability. We now consider whether relative predictability is correlated with a larger constituent’s accent pattern. The following analysis focuses on accent patterns of non-trivial base NPs.<sup>3</sup> For this study we labeled base NPs by hand for the corpora described in Section 2. For each base NP, we calculate which word is the most predictable and which is the least. We want to see, when comparing with its neighboring

<sup>2</sup>For example, in a 7.5M word general news corpus (from CNN and Reuters), *street* occurs 2115 times and *avenue* just 194. Therefore, the unigram predictability of *street* is higher than that of *avenue*. The most common bigram with *street* is *Wall Street* which occurs 116 times and the most common bigram with *avenue* is *Pennsylvania Avenue* which occurs 97. In this domain, the bigram predictability for *street* in *Fifth Street* is extremely low because this combination never occurred, while that for *avenue* in *Fifth Avenue* is -3.0995 which is the third most predictable bigrams with *avenue* as the second word.

<sup>3</sup>Non-recursive noun phrases containing at least two elements.

Corpus	Predictability Model	Performance	Standard Error
Read	baseline model	81.98%	
	unigram model	82.86%	$\pm 0.93$
	bigram predictability model	84.41%	$\pm 1.10$
	unigram+bigram model	85.03%	$\pm 1.04$
Spontaneous	baseline model	70.03%	
	unigram model	72.22%	$\pm 0.62$
	bigram model	74.46%	$\pm 0.30$
	unigram+bigram model	77.43%	$\pm 0.51$

Table 4: Ripper Results for Accent Status Prediction

Model	Predictability	Total	Accented Word	Not Accented	Accentability
unigram	Least Predictable	1206	877	329	72.72%
	Most Predictable	1198	485	713	40.48%
bigram	Least Predictable	1205	965	240	80.08%
	Most Predictable	1194	488	706	40.87%

Table 5: Relative Predictability and Accent Status

words, whether the most predictable word is more likely to be deaccented. As shown in Table 5, the “total” column represents the total number of most (or least) predictable words in all baseNPs<sup>4</sup>. The next two columns indicate how many of them are accented and deaccented. The last column is the percentage of words that are accented. Table 5 shows that the probability of accenting a most predictable word is between 40.48% and 45.96% and that of a least predictable word is between 72.72% and 80.08%. This result indicates that relative predictability is also a useful predictor for a word’s accentability.

## 8 Discussion

It is difficult to directly compare our results with previous accent prediction studies, to determine the general utility of bigram predictability in accent assignment, due to differences in domain and the scope of our task. For example, Hirschberg (1993) built a comprehensive accent prediction model using machine learning techniques for predicting accent status for all word classes for a text-to-speech system, employing part-of-speech, various types of information status inferred from the text, and a number of distance metrics, as well as a complex nominal predictor developed by Sproat (1992). An algorithm making use of these features achieved 76.5%-80% accent prediction accuracy for a broadcast news

corpus, 85% for sentences from the ATIS corpus of spontaneous elicited speech, and 98.3% success on a corpus of laboratory read sentences. Liberman and Sproat’s (1992) success in predicting accent patterns for complex nominals alone, using rules combining a number of features, achieved considerably higher success rates (91% correct, 5.4% acceptable, 3.6% unacceptable when rated by human subjects) for 500 complex nominals of 2 or more elements chosen from the AP Newswire. Our results, using bigram predictability alone, 77% for the spontaneous corpus and 85% for the read corpus, and using a different success estimate, while not as impressive as (Liberman and Sproat, 1992)’s, nonetheless demonstrate the utility of a relatively untested feature for this task.

In this paper, we have investigated several collocation-based measures for pitch accent prediction. Our initial hypothesis was that word collocation affects pitch accent placement, and that the more predictable a word is in terms of its local lexical context, the more likely it is to be deaccented. In order to verify this claim, we estimated three collocation measures: word predictability, mutual information and the Dice coefficient. We then used statistical techniques to analyze the correlation between our different word collocation metrics and pitch accent assignment for nouns. Our results show that, of all the collocation measures we investigated, bigram word predictability has the strongest correla-

<sup>4</sup>The total number of most predictable words is not equal to that of least predictable words due to ties.

tion with pitch accent assignment. Based on this finding, we built several pitch accent models, assessing the usefulness of unigram and bigram word predictability—as well as a combined model—in accent predication. Our results show that the bigram model performs consistently better than the unigram model, which does not incorporate local context information. However, our combined model performs best of all, suggesting that both contextual and non-contextual features of a word are important in determining whether or not it should be accented.

These results are particularly important for the development of future accent assignment algorithms for text-to-speech. For our continuing research, we will focus on two directions. The first is to combine our word predictability feature with other pitch accent predictors that have been previously used for automatic accent prediction. Features such as information status, grammatical function, and part-of-speech, have also been shown to be important determinants of accent assignment. So, our final pitch accent model should include many other features. Second, we hope to test whether the utility of bigram predictability can be generalized across different domains. For this purpose, we have collected an annotated AP news speech corpus and an AP news text corpus, and we will carry out a similar experiment in this domain.

## 9 Acknowledgments

Thanks for C. Jin, K. McKeown, R. Barzilay, J. Shaw, N. Elhadad, M. Kan, D. Jordan, and anonymous reviewers for the help on data preparation and useful comments. This research is supported in part by the NSF Grant IRI 9528998, the NLM Grant R01 LM06593-01 and the Columbia University Center for Advanced Technology in High Performance Computing and Communications in Healthcare.

## References

- D. Bolinger. 1961. Contrastive accent and contrastive stress. *language*, 37:83–96.
- D. Bolinger. 1989. *Intonation and Its Uses*. Stanford University Press.
- G. Brown. 1983. Prosodic structure and the given/new distinction. In A. Cutler and D.R. Ladd, ed., *Prosody: Models and Measurements*, pages 67–78. Springer-Verlag, Berlin.
- P. Cohen. 1995a. *Empirical methods for artificial intelligence*. MIT press, Cambridge, MA.
- W. Cohen. 1995b. Fast effective rule induction. In *Proc. of the 12th International Conference on Machine Learning*.
- W. J. Conover. 1980. *Practical Nonparametric Statistics*. Wiley, New York, 2nd edition.
- M. Dalal, S. Feiner, K. McKeown, S. Pan, M. Zhou, T. Hoellerer, J. Shaw, Y. Feng, and J. Fromer. 1996. Negotiation for automated generation of temporal multimedia presentations. In *Proc. of ACM Multimedia 96*, pages 55–64.
- Lee R. Dice. 1945. Measures of the amount of ecologic association between species. *Journal of Ecology*, 26:297–302.
- Robert M. Fano. 1961. *Transmission of Information: A Statistical Theory of Communications*. MIT Press, Cambridge, MA.
- J. Hirschberg. 1993. Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence*, 63:305–340.
- D. Robert Ladd. 1996. *Intonational Phonology*. Cambridge University Press, Cambridge.
- M. Liberman and R. Sproat. 1992. The stress and structure of modified noun phrases in English. In I. Sag, ed., *Lexical Matters*, pages 131–182. University of Chicago Press.
- J. Marchand. 1993. Message posted on HUMANIST mailing list, April.
- C. Nakatani. 1998. Constituent-based accent prediction. In *Proc. of COLING/ACL '98*, pages 939–945, Montreal, Canada.
- S. Pan and K. McKeown. 1998. Learning intonation rules for concept to speech generation. In *Proc. of COLING/ACL '98*, Montreal, Canada.
- S. Pan and K. McKeown. 1999. Word informativeness and automatic pitch accent modeling. In *Proc. of the Joint SIGDAT Conference on EMNLP and VLC*, pages 148–157.
- K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 1992. ToBI: a standard for labeling English prosody. In *Proc. of ICSLP92*.
- R. Sproat, J. Hirschberg, and D. Yarowsky. 1992. A corpus-based synthesizer. In *Proc. of ICSLP92*, pages 563–566, Banff.
- R. Sproat, ed. 1998. *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*. Kluwer.