Word Informativeness and Automatic Pitch Accent Modeling

Shimei Pan and **Kathleen R. McKeown**
Dept. of Computer Science
Columbia University
New York, NY 10027, USA
{pan,kathy}@cs.columbia.edu

To appear in Proc. of EMNLP/VLC, 1999.

Abstract

In intonational phonology and speech synthesis research, it has been suggested that the relative informativeness of a word can be used to predict pitch prominence. The more information conveyed by a word, the more likely it will be accented. But there are others who express doubts about such a correlation. In this paper, we provide some empirical evidence to support the existence of such a correlation by employing two widely accepted measures of informativeness. Our experiments show that there is a positive correlation between the informativeness of a word and its pitch accent assignment. They also show that informativeness enables statistically significant improvements in pitch accent prediction. The computation of word informativeness is inexpensive and can be incorporated into speech synthesis systems easily.

1 Introduction

The production of natural, intelligible speech remains a major challenge for speech synthesis research. Recent research has focused on prosody modeling (Silverman, 1987; Hirschberg, 1990; Santen, 1992), which determines the variations in pitch, tempo and rhythm. One of the critical issues in prosody modeling is pitch accent assignment. Pitch accent is associated with the pitch prominence of a word. For example, some words may sound more prominent than others within a sentence because they are associated with a sharp pitch rise or fall. Usually, the prominent words bear pitch accents while the less prominent ones do not. Although native speakers of a particular language have no difficulty in deciding which words in their utterances should be accented, the general pattern of accenting in a language, such as English, is still an open question.

Some linguists speculate that relative informativeness, or semantic weight of a word can influence accent placement. Ladd (1996) claims that "the speakers assess the relative semantic weight or informativeness of potentially accentable words and put the accent on the most informative point or points" (ibid, pg. 175). He also claims that "if we understand relative semantic weight, we will automatically understand accent placement" (ibid, pg. 186). Bolinger (Bolinger, 1972) also uses the following examples to illustrate the phenomenon:

1. "*He was arrested because he KILLED a man.*"
2. "*He was arrested because he killed a POLICEMAN.*"

The capitalized words in the examples are accented. In (1), "*man*" is semantically empty relative to "*kill*"; therefore, the verb "*kill*" gets accented. However, in (2), "*policeman*" is semantically rich and is accented instead.

However, different theories, not based on informativeness, were proposed to explain the above phenomenon. For example, Bresnan's (1971) explanation is based on syntactic function. She suggests that "*man*" in the above sentence does not get accented because "*man*" and other words like "*guy*" or "*person*" or "*thing*" form a category of "semi-pronouns". Counter-examples listed below raise more questions about the usefulness of semantic informativeness. The accent pattern in the following examples cannot be explained solely by semantic informativeness.

3. "*HOOVER dam.*"
4. "*Hoover TOWER.*"

While researchers have discussed the possible influence of semantic informativeness, there has been no known empirical study of the claim nor has this type of information been incorporated into computational models of prosody. In this work, we employ two measurements of informativeness. First, we adopt an information-based framework (Shannon, 1948), quantifying the "Information Content (IC)" of a word as the negative log likelihood of a word in a corpus. The second measurement is TF*IDF (Term Frequency times Inverse Document Frequency) (Salton, 1989; Salton, 1991), which has been widely used to quantify word importance in information retrieval tasks. Both IC and TF*IDF are well established measurements of informativeness and therefore, good candidates to investigate. Our empirical study shows that word informativeness not only is closely related to word accentuation, but also provides new power in pitch accent prediction. Our results suggest that information content is a valuable feature to be incorporated in speech synthesis systems.

In the following sections, we first define IC and TF*IDF. Then, a description of the corpus used in this study is provided. We then describe a set of experiments conducted to study the relation between informativeness and pitch accent. We explain how machine learning techniques are used in the pitch accent modeling process. Our results show that:

- Both IC and TF*IDF scores are strongly correlated with pitch accent assignment.
- IC is a more powerful predictor than TF*IDF.
- IC provides better prediction power in pitch accent prediction than previous techniques.

The investigated pitch accent models can be easily adopted by speech synthesis systems.

2 Definitions of IC and TF*IDF

Following the standard definition in information theory (Shannon, 1948; Fano, 1961; Cover and Thomas, 1991) the IC of a word is

$$IC(w) = -log(P(w))$$

where $P(w)$ is the probability of the word w appearing in a corpus and $P(w)$ is estimated as: $\frac{F(w)}{N}$ where $F(w)$ is the frequency of w in the corpus and N is the accumulative occurrence of all the words in the corpus. Intuitively, if the probability of a word increases, its informativeness decreases and therefore it is less likely to be an information focus. Similarly, it is therefore less likely to be communicated with pitch prominence.

TF*IDF is defined by two components multiplied together. TF (Term Frequency) is the word frequency within a document; IDF (Inverse Document Frequency) is the logarithm of the ratio of the total number of documents to the number of documents containing the word. The product of TF*IDF is higher if a word has a high frequency within the document, which signifies high importance for the current document, and low dispersion in the corpus, which signifies high specificity. In this research, we employed a variant of TF*IDF score used in SMART (Buckley, 1985), a popular information retrieval package:

$$(\text{TF*IDF})_{w_i,d_j} = \frac{(1.0 + \log F_{w_i,d_j}) \log \frac{N}{N_{w_i}}}{\sqrt{\sum_{k=1}^{M} ((1.0 + \log F_{w_k,d_j}) \log \frac{N}{N_{w_k}})^2}}$$

where F_{w_i,d_j} is the the frequency of word w_i in document d_j, N is the total number of documents, N_{w_i} is the number of documents containing word w_i and M is the number of distinct stemmed words in document d_j.

IC and TF*IDF capture different kinds of informativeness. IC is a matrix global in the domain of a corpus and each word in a corpus has a unique IC score. TF*IDF captures the balance of a matrix local to a given document (TF) and a matrix global in a corpus (IDF). Therefore, the TF*IDF score of a word changes from one document to another (different TF). However, some global features are also captured by TF*IDF. For example, a common word in the domain tends to get low TF*IDF score in all the documents in the corpus.

3 Corpus Description

In order to empirically study the relations between word informativeness and pitch accent, we use a medical corpus which includes a speech portion and a text portion. The speech corpus includes fourteen segments which total about 30 minutes of speech. The speech was collected at Columbia Presbyterian Medical Center (CPMC) where doctors informed residents or nurses about the post-operative status of a patient who has just undergone a bypass surgery. The speech corpus was transcribed orthographically by a medical professional and is also intonationally labeled with pitch accents by a ToBI (Tone and Break Index) (Silverman et al., 1992; Beckman and Hirschberg, 1994) expert. The text corpus includes 1.24 million, 2,422 discharge summaries, spanning a larger group of patients. The majority of the patients have also undergone cardiac surgery. The orthographic transcripts as well as the text corpus are used to calculate the IC and TF*IDF scores. First, all the words in the text corpus as well as the speech transcripts are processed by a stemming model so that words like "*receive*" and "*receives*" are treated as one word. We employ a revised version of Lovins' stemming algorithm (Lovins, 1968) which is implemented in SMART. Although the usefulness of stemming is arguable, we choose to use stemming because we think "*receive*" and "*receives*" are equally likely to be accented. Then, IC and TF*IDF are calculated. After this, the effectiveness of informativeness in accent placement is verified using the speech corpus. Each word in the speech corpus has an IC score, a TF*IDF score, a part-of-speech (POS) tag and a pitch accent label. Both IC and TF*IDF are used to test the correlation between informativeness and accentuation. POS is also investigated by several machine learning techniques in automatic pitch accent modeling.

4 Experiments

We conducted a series of experiments to determine whether there is a correlation between informativeness and pitch accent and whether informativeness provides an improvement over other known indicators on pitch accent, such as part-of-speech. We experimented with different forms of machine learning to integrate indicators within a single framework, testing whether rule induction or hidden Markov modeling provides a better model.

4.1 Ranking Word Informativeness in the Corpus

Table 1 and 2 shows the most and least informative words in the corpus. The IC order indicates the rank among all the words in the corpus, while TF*IDF order in the table indicates the rank among the words within a document. The document was picked randomly from the corpus. In general, most of the least informative words are function words, such as "*with*" or "*and*". However, some content words are selected, such as "*patient*", "*year*", "*old*". These content words are very common in this domain and are mentioned in almost all the documents in the corpus. In contrast, the majority of the most informative words are content words. Some of the selections are less expected. For example "*your*" ranks as the most informative word in a document using TF*IDF. This indicates that listeners or readers are rarely addressed in the corpus. It appears only once in the entire corpus.

Rank	IC Most Informative		IC Least Informative	
	Words	IC	Words	IC
1	zophrin	14.02725	with	4.08777
2	name1	14.02725	on	4.20878
3	xyphoid	14.02725	patient	4.26354
4	wytensin	14.02725	in	4.35834
5	pyonephritis	14.02725	she	4.52409
6	orobuccal	14.02725	he	4.52918
7	tzanck	14.02725	for	4.66436
8	synthetic	14.02725	no	4.69019
9	Rx	14.02725	day	4.78832
10	quote	14.02725	had	4.98343

Table 1: IC Most and Least informative words

Rank	TF*IDF Most Informative		TF*IDF Least Informative	
	Words	TF*IDF	Words	TF*IDF
1	your	0.15746	and	0.00008
2	vol	0.15238	a	0.00009
3	tank	0.15238	the	0.00009
4	sonometer	0.15238	to	0.00016
5	papillary	0.15238	was	0.00020
6	pancuronium	0.15238	of	0.00024
7	name2	0.15238	with	0.00034
8	name3	0.15238	in	0.00041
9	incomplete	0.14345	old	0.00068
10	yes	0.13883	year	0.00088

Table 2: TF*IDF Most and Least informative words

4.2 Testing the Correlation of Informativeness and Accent Prediction

In order to verify whether word informativeness is correlated with pitch accent, we employ Spearman's rank correlation coefficient ρ and associated test (Conover, 1980) to estimate the correlations between IC and pitch prominence as well as TF*IDF and pitch prominence. As shown in Table 3, both IC and TF*IDF are closely correlated to pitch accent with a significance level $p = 2.67 \cdot 10^{-65}$ and $p = 2.90 \cdot 10^{-84}$ respectively. Because the correlation coefficient ρ is positive, this indicates that the higher the IC and TF*IDF are, the more likely a word is to be accented.

4.3 Learning IC and TF*IDF Accent Models

The correlation test suggests that there is a strong connection between informativeness and pitch accent. But we also want to show how much performance gain can be achieved by adding this information to pitch accent models. To study the effect of TF*IDF and IC on pitch accent, we use machine learning techniques to learn models that predict the

Feature	Correlation Coefficient	Significance Level
TF*IDF	$\rho = 0.29$	$p = 2.67 \cdot 10^{-65}$
IC	$\rho = 0.34$	$p = 2.90 \cdot 10^{-84}$

Table 3: The Correlation of Informativeness and Accentuation

effect of these indicators on pitch accent. We use both RIPPER (Cohen, 1995) and Hidden Markov Models (HMM) (Rabiner and Juang, 1986) to build pitch accent models. RIPPER is a system that learns sets of classification rules from training data. It automatically selects rules which maximize the information gain and employs heuristics to decide when to stop to prevent over-fitting. The performance of RIPPER is comparable with most benchmark rule induction systems such as C4.5 (Quinlan, 1993). We train RIPPER on the speech corpus using 10-fold cross-validation, a standard procedure for training and testing when the amount of data is limited. In this experiment, the predictors are IC or TF*IDF, and the response variable is the pitch accent assignment. Once a set of RIPPER rules are acquired, they can be used to predict which word should be accented in a new corpus.

HMM is a probability model which has been successfully used in many applications, such as speech recognition (Rabiner, 1989) and part-of-speech tagging (Kupiec, 1992). A HMM is defined as a triple: λ=(A, B, Π) where A is a state transition probability matrix, B is a observation probability distribution matrix, and Π is an initial state distribution vector. In this experiment, the hidden states are the accent status of words which can be either “accented” or “not accented”. The observations are IC or TF*IDF score of each word. Because of the limitation of the size of the speech corpus, we use a first-order HMM where the following condition is assumed:

$$P(Q_{t+1}=i|Q_t=j, Q_{t-1}=k, \ldots Q_1=n) = P(Q_{t+1}=i|Q_t=j)$$

where Q_t is the state at time t. Because we employ a supervised training process, no sophisticated parameter estimation procedure, such as the Baum-Welch algorithm (Rabiner, 1989) is necessary. Here all the parameters are precisely calculated using the following formula:

$$A = \{\alpha_{ij} : i = 1, .., N, j = 1, .., N\}$$

$$\alpha_{ij} = \frac{F(Q_{t-1} = i, Q_t = j)}{F(Q_t = j)}$$

$$B = \{\beta_{jm} : j = 1, .., N, m = 1, .., M\}$$

$$\beta_{jm} = \frac{F(Q_t = j, O_t = m)}{F(Q_t = j)}$$

$$\Pi = \{\pi_i : i = 1, .., N\}$$

$$\pi_i = \frac{F(Q_1 = i)}{F(Q_1)}$$

where N is the number of hidden states and M is the number of observations.

Once all the parameters of a HMM are set, we employ the Viterbi algorithm (Viterbi, 1967; Forney, 1973) to find an optimal accentuation sequence which maximizes the possibility of the occurrence of the observed IC or TF*IDF sequence given the HMM.

Both RIPPER and HMM are widely accepted machine learning systems. However, their theoretical bases are very different. HMM focuses on optimizing a sequence of accent assignments instead of isolated accent assignment. By employing both of them, we want to show that our conclusions hold for both approaches. Furthermore, we expect HMM to do better than RIPPER because the influence of context words is incorporated.

We use a baseline model where all words are assigned a default accent status (accented). 52% of the words in the corpus are actually accented and thus, the baseline has a performance of 52%. Our results in

Models	HMM Performance	RIPPER Performance
Baseline	52.02%	52.02%
TF*IDF Model	67.25%	65.66%
IC Model	71.96%	70.06%

Table 4: Comparison of IC, TF*IDF model with the baseline model

Table 4 show that when TF*IDF is used to predict pitch accent, performance is increased over the baseline of 52% to 67.25% and 65.66 % for HMM and RIPPER respectively. In the IC model, the performance is further increased to 71.96% and 70.06%. These results are obtained by using 10-fold cross-validation. We can draw two conclusions from the results. First, both IC and TF*IDF are very effective in pitch accent prediction. All the improvements over the baseline model are statistically significant with $p < 1.11 \cdot 10^{-16}$ [1], using χ^2 test (Fienberg, 1983; Fleiss, 1981). Second, the IC model is more powerful than the TF*IDF model. It out performs the TF*IDF model with $p = 3.8 \cdot 10^{-5}$ for the HMM model and $p = 0.0002$ for the RIPPER model. The low p-values show the improvements achieved by the IC models are significant. Since IC performs better than TF*IDF in pitch accent prediction, we choose IC to measure informativeness in all the following experiments. Another observation of the results is that the HMM models do show some improvements over the RIPPER models. But the difference is marginal. More data is needed to test the significance of the improvements.

4.4 Incorporating IC in Reference Accent Models

In order to show that IC provides additional power in predicting pitch accent than current models, we need to directly compare the influence of IC with that of other reference models. In this section, we describe experiments that compare IC alone against a part-of-speech (POS) model for pitch accent prediction and then compare a model that integrates IC with POS against the POS model. Finally, anticipating the possibility that other features within a traditional TTS in combination with POS may provide equal or better performance than the addition of IC, we carried out experiments that directly compare the performance of Text-to-Speech (TTS) synthesizer alone with a model that integrates TTS with IC.

In most speech synthesis systems, part-of-speech (POS) is the most powerful feature in pitch accent prediction. Therefore, showing that IC provides additional power over POS is important. In addition to the importance of POS within TTS for predicting pitch accent, there is a clear overlap between POS and IC. We have shown that the words with highest IC usually are content words and the words with lowest IC are frequently function words. This is an added incentive for comparing IC with POS models. Thus, we want to explore whether the new information added by IC can provide any improvement when both of them are used to predict accent assignment.

In order to create a POS model, we first utilize MXPOST, a maximum entropy part-of-speech tagger (Ratnaparkhi, 1996) to get the POS information for each word. The performance of the MXPOST tagger is comparable with most benchmark POS taggers, such as Brill's tagger (Brill, 1994). After this, we map all the part-of-speech tags into seven categories: "noun", "verb", "adjective", "adverb", "number", "pronoun" and "others". The mapping procedure is conducted because keeping all the initial tags (about 35) will drastically increase the requirements for the amount of training data.

[1] S reports p=0 because of underflow. The real p value is less than $1.11 \cdot 10^{-16}$, which is the smallest value the computer can represent in this case

Models	HMM Performance	RIPPER Performance
IC Model	71.96%	70.06%
POS Model	71.33%	70.52%
POS+IC Model	74.06%	73.71%

Table 5: Comparison of POS+IC model with POS model

Models	HMM Performance	RIPPER Performance
TTS Model	71.75%	71.75%
TTS+IC Model	72.30%	72.75%
POS+IC Model	74.06%	73.71%

Table 6: Comparison of TTS+IC model with TTS model

The obtained POS tag is the predictor in the POS model. As shown in table 5, the performance of these two POS models are 71.33% and 70.52% for HMM and RIPPER respectively, which is comparable with that of the IC model. This comparison further shows the strength of IC because it has similar power to POS in pitch accent prediction and it is very easy to compute. When the POS models are augmented with IC, the POS+IC model performance is increased to 74.06% and 73.71% respectively. The improvement is statistically significant with $p = 0.015$ for HMM model and $p = 0.005$ for RIPPER which means the new information captured by IC provides additional predicting power for the POS+IC models. These experiments produce new evidence confirming that IC is a valuable feature in pitch accent modeling.

We also tried another reference model, Text-to-Speech (TTS) synthesizer output, to evaluate the results. The TTS pitch accent model is more comprehensive than the POS model. It has taken many features into consideration, such as discourse and semantic information. It is well established and has been evaluated in various situations. In this research, we adopted Bell Laboratories' TTS system (Sproat, 1997; Olive and Liberman, 1985; Hirschberg, 1990). We run it on our corpus first to get the TTS pitch accent assignments. Comparing the TTS accent assignment with the expert accent assignment, the TTS performance is 71.75% which is statistically significantly lower than the HMM POS+IC model with $p = 0.039$. We also tried to incorporate IC in TTS model. A simple way of doing this is to use the TTS output and IC as predictors and train them with our data. The obtained TTS+IC models achieve marginal improvement. The performance of TTS+IC model increases to 72.30% and 72.75% for HMM and RIPPER respectively, which is lower than that of the POS+IC models. We speculate that this is may be due to the corpus we used. The Bell Laboratories' TTS pitch accent model is trained in a totally different domain, and our medical corpus seems to negatively affect the TTS performance (71.75% compared to around 80%, its normal performance). Since the TTS+IC models involve two totally different domains, the effectiveness of IC may be compromised. If this assumption holds, we think that the TTS+IC model will perform better when IC is trained together with the TTS internal features on our corpus directly. But since this requires retraining a TTS system for a new domain and it is very hard for us to conduct such an experiment, no further comparison was conducted to verify this assumption.

Although TF*IDF is less powerful than IC in pitch accent prediction, since they measure two different kinds of informativeness, it is possible that a TF*IDF+IC model can perform better than the IC model. Similarly, if TF*IDF is incorporated in the POS+IC model, the overall performance may increase for the POS+IC+TF*IDF model. However, our experiment shows no improvements when TF*IDF is incorporated in the IC and POS+IC model. Our experiments show that IC is always the dominant predictor when both IC and TF*IDF are presented.

5 Related Work

Information based approaches were applied in some natural languages applications before. In (Resnik, 1993; Resnik, 1995), IC was used to measure semantic similarity between words and it is shown to be more effective than traditional measurements of semantic distance within the WordNet hierarchy. A similar log-based information-like measurement was also employed in (Leacock and Chodorow, 1998) to measure semantic similarity. TF*IDF scores are mainly used in keyword-based information retrieval tasks. For example, TF*IDF has been used in (Salton, 1989; Salton, 1991) to index the words in a document and is also implemented in SMART (Buckley, 1985) which is a general-purpose information retrieval package, providing basic tools and libraries to facilitate information retrieval tasks.

Some early work on pitch accent prediction in speech synthesis only uses the distinction between content words and function words. Although this approach is simple, it tends to assign more pitch accents than necessary. We also tried the content/function word model on our corpus and as expected, we found it to be less powerful than the part-of-speech model. More advanced pitch accent models make use of other information, such as part-of-speech, given/new distinctions and contrast information (Hirschberg, 1993). Semantic information is also employed in predicting accent patterns for complex nominal phrases (Sproat, 1994). Other comprehensive pitch accent models have been suggested in (Pan and McKeown, 1998) in the framework of Concept-to-Speech generation where the output of a natural language generation system is used to predict pitch accent.

6 Discussion

Since IC is not a perfect measurement of informativeness, it can cause problems in accent prediction. Moreover, even if a perfect measurement of informativeness is available, more features may be needed in order to build a satisfactory pitch accent model. In this section, we discuss each of these issues.

IC does not directly measure the informativeness of a word. It measures the rarity of a word in a corpus. That a word is rare doesn't necessarily mean that it is informative. Semantically empty words can be ranked high using IC as well. For example, CABG is a common operation in this domain. "*CABG*" is almost always used whenever the operation is mentioned. However, in a few instances, it is referred to as a "*CABG operation*". As a result, the semantically empty word (in this context) "*operation*" gets a high IC score and it is very hard to distinguish high IC scores resulting from this situation from those that accurately measure informativeness and this causes problems in precisely measuring the IC of a word. Similarly, misspelled words also can have high IC score due to their rarity.

Although IC is not ideal for quantifying word informativeness, even with a perfect measurement of informativeness, there are still many cases where this information by itself would not be enough. For example, each word only gets a unique IC score regardless of its context; yet it is well known that context information, such as given/new and contrast, plays an important role in accentuation. In the future, we plan to build a comprehensive accent model with more pitch accent indicators, such as syntactic, semantic and discourse features.

7 Conclusion

In this paper, we have provided empirical evidence for the usefulness of informativeness for accent assignment. Overall, there is a positive correlation between indicators of informativeness, such as IC and TF*IDF, and pitch accent. The more informative a word is, the more likely that a pitch accent is assigned to the word. Both of the two measurements of informativeness improve over the baseline performance significantly. We also show that IC is a more powerful measure of informativeness than TF*IDF for pitch accent prediction. Later, when comparing IC-empowered POS models with POS models, we found that IC enables additional, statistically significant improvements for pitch accent assignment. This performance also outperforms the TTS pitch accent model significantly. Overall, IC is not only effective, as shown in the results, but also relatively inexpensive to compute for a new domain. Almost all speech synthesis systems, text-to-speech as well as concept-to-speech systems, can employ this feature as long as there is a large corpus. In the future, we plan to explore other information content measurements and incorporate them in a more comprehensive accent model with more discourse and semantic features included.

8 Acknowledgement

Thanks to Julia Hirschberg, Vasileios Hatzivassiloglou and James Shaw for comments and suggestions on an earlier version of this paper. Thanks to Desmand Jordan for helping us with the collection of the speech and text corpus. This research is supported in part by the National Science Foundation under Grant No. IRI 9528998, the National Library of Medicine under project R01 LM06593-01 and the Columbia University Center for Advanced Technology in High Performance Computing and Communications in Healthcare (funded by the New York State Science and Technology Foundation).

References

Mary Beckman and Julia Hirschberg. 1994. The ToBI annotation conventions. Technical report, Ohio State University, Columbus.

Dwight Bolinger. 1972. Accent is predictable (if you're a mind-reader). *Language*, 48:633–644.

Joan Bresnan. 1971. Sentence stress and syntactic transformations. *Language*, 47:257–280.

Eric Brill. 1994. Some advances in rule-based part of speech tagging. In *Proceedings of the 12th National Conference on Artificial Intelligence*.

Chris Buckley. 1985. Implementation of the SMART information retreival system. Technical Report 85-686, Cornell University.

William Cohen. 1995. Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning*.

W. J. Conover. 1980. *Practical Nonparametic Statistics*. Wiley, New York, 2nd edition.

Thomas M. Cover and Joy A. Thomas. 1991. *Elements of Information Theory*. Wiley, New York.

Robert M. Fano. 1961. *Transmission of Information: A Statistical Theory of Communications*. MIT Press, Cambridge, Massachusetts.

Stephen E. Fienberg. 1983. *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, Mass, 2nd edition.

Joseph L. Fleiss. 1981. *Statistical Methods for Rates and Proportions*. Wiley, New York, 2nd edition.

G. David Forney. 1973. The Viterbi algorithm. *Proceedings of IEEE*, 61(3).

Julia Hirschberg. 1990. Assigning pitch accent in synthetic speech: The given/new distinction and deaccentability. In *Proceedings of the Seventh National Conference of American Association of Artificial Intelligence*, pages 952–957, Boston.

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

Julian Kupiec. 1992. Robust part-of-speech tagging using a hidden markov model.

Computer Speech and Language, 6(3):225–242, July.

D. Robert Ladd. 1996. *Intonational Phonology*. Cambridge University Press, Cambridge.

Claudia Leacock and Martin Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database*, chapter 11. MIT Press.

Julie Beth Lovins. 1968. Development of a stemming algorithm. In *Mechanical Translation and Computational Linguistics*, volume 11.

Joseph. P. Olive and Mark Y. Liberman. 1985. Text to Speech—An overview. *Journal of the Acoustic Society of America*, 78(Fall):s6.

Shimei Pan and Kathleen R. McKeown. 1998. Learning intonation rules for concept to speech generation. In *Proceedings of COLING/ACL'98*, Montreal, Canada.

John R. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo.

Lawrence R. Rabiner and B. H. Juang. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–15, January.

Lawrence R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Adwait Ratnaparkhi. 1996. A maximum entropy part of speech tagger. In Eric Brill and Kenneth Church, editors, *Conference on Empirical Natural Language Processing*. Univ. of Pennsylvania.

Philip Resnik. 1993. Semantic classes and syntactic ambiguity. In *Proc. of ARPA Workshop on Human Language Technology*, pages 278–283. Morgan Kaufmann Publishers.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, Montreal, Canada.

Gerard Salton. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, Massachusetts.

Gerard Salton. 1991. Developments in automatic text retrieval. *Science*, 253:974–980, August.

Jan P. H. Van Santen. 1992. Contextual effects on vowel duration. *Speech Communication*, 11:513–546, January.

Claude E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October.

Kim Silverman, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. ToBI: a standard for labelling English prosody. In *Proceedings of ICSLP92*, volume 2.

Kim Silverman. 1987. *The structure and processing of fundamental frequency contours*. Ph.D. thesis, Cambridge University.

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

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

Andrew J. Viterbi. 1967. Error bound for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions in Information Theory*, 13(2).