To Memorize or to Predict: Prominence Labeling in Conversational Speech

A. Nenkova, J. Brenier, A. Kothari, S. Calhoun[†], L. Whitton, D. Beaver, D. Jurafsky
Stanford University
{anenkova,jbrenier,anubha,lwhitton,dib,jurafsky}@stanford.edu
†University of Edinburgh
Sasha.Calhoun@ed.ac.uk

Abstract

The immense prosodic variation of natural conversational speech makes it challenging to predict which words are prosodically *prominent* in this genre. In this paper, we examine a new feature, *accent ratio*, which captures how likely it is that a word will be realized as prominent or not. We compare this feature with traditional accent-prediction features (based on part of speech and N-grams) as well as with several linguistically motivated and manually labeled information structure features, such as whether a word is *given*, *new*, or *contrastive*. Our results show that the linguistic features do not lead to significant improvements, while accent ratio alone can yield prediction performance almost as good as the combination of any other subset of features. Moreover, this feature is useful even across genres; an accent-ratio classifier trained only on conversational speech predicts prominence with high accuracy in broadcast news. Our results suggest that carefully chosen lexicalized features can outperform less fine-grained features.

1 Introduction

Being able to predict the *prominence* or *pitch accent* status of a word in conversational speech is important for implementing text-to-speech in dialog systems, as well as in detection of prosody in conversational speech recognition.

Previous investigations of prominence prediction from text have primarily relied on robust surface features with some deeper information structure features. Surface features like a word's part-of-speech (POS) (Hirschberg, 1993) and its unigram and bigram probability (Pan and McKeown, 1999; Pan and Hirschberg, 2000) are quite useful; content words are much more likely to be accented than function words, and words with higher probability are less likely to be prominent. More sophisticated linguistic features have also been used, generally based on information-structural notions of *contrast*, *focus*, or *given-new*. (Hirschberg, 1993).

For example, in the Switchboard utterance below, there is an intrinsic contrast between the words "women" and "men", making both terms more salient (words in all capital letters represent prominent tokens):

you SEE WOMEN$_c$ GOING off to WARS as WELL as MEN$_c$.

Similarly the givenness of a word may help determine its prominence. The speaker needs to focus the hearer's attention on *new* entities in the discourse, so these are likely to be realized as prominent. *Old* entities, on the other had, need not be prominent; these tendencies can be seen in the following example.

they$_{old}$ have all the WATER$_{new}$ they$_{old}$ WANT. they$_{old}$ can ACTUALLY PUMP water$_{old}$.

While previous models have attempted to capture global properties of words (via POS or unigram probability), they have not in general used word identity as a predictive feature, assuming either that current supervised training sets would be too small or that word identity would not be robust across genres (Pan et al., 2002). In this paper, we show a way to capture word identity in a feature, **accent ratio**, that works well with current small supervised training sets, and is robust to genre differences.

We also use a corpus which has been hand-labeled for information structure features (including given/new and contrast information) to investigate the relative usefulness of both linguistic and shallow features, as well as how well different features combine with each other.

[0] Thanks to the Edinburgh-Stanford Link and ONR (MURI award N000140510388) for generous support.

2 Data and features

For our experiments we use 12 Switchboard (Godfrey et al., 1992) conversations, 14,555 tokens in total. Each word was manually labeled for presence or absence of pitch accent[1] , as well as additional features including information status (or givenness), contrast and animacy distinctions, (Nissim et al., 2004; Calhoun et al., 2005; Zaenen et al., 2004), features that linguistic literature suggests are predictive of prominence (Bolinger, 1961; Chafe, 1976).

All of the features described in detail below have been shown to have statistically significant correlation with prominence (Brenier et al., 2006).

Information status The information status (IS), or givenness, of discourse entities is important for choosing appropriate reference form (Prince, 1992; Gundel et al., 1993) and possibly plays a role in prominence decisions as well (Brown, 1983). No previous studies have examined the usefulness of information status in *naturally occurring* conversational speech. The annotation in our corpus is based on the givenness hierarchy of Prince: first mentions of entities were marked as *new* and subsequent mentions as *old*. Entities that are not previously mentioned, but that are generally known or semantically related to other entities in the preceding context are marked as *med*iated. Obviously, the givenness annotation applies only to referring expressions, i.e. noun phrases the semantic interpretation of which is a discourse entity. This restriction inherently limits the power of the feature for prominence prediction, which has to be performed for all classes of words. Complete details of the IS annotation can be found in (Nissim et al., 2004).

Kontrast One reason speakers make entities in an utterance prominence is because of information structure considerations (Rooth, 1992; Vallduví and Vilkuna, 1998). That is, parts of an utterance which distinguish the information the speaker actually says from the information they could have said, are made salient, e.g. because that information answers a question, or contrasts with a similar entity in the context. Several possible triggers of this sort of salience were marked in the corpus, with words that were not kontrastive (in this sense) being marked as *background*:

- *contrastive* if the word is directly differentiated from a previous topical or semantically-related word;
- *subset* if it refers to a member of a more general set mentioned in the surrounding context;
- *adverbial* if a focus-sensitive adverb such as "only" or "even" is associated with the word being annotated;
- *correction* if the speaker intended to correct or clarify a previous word or phrase;
- *answer* if the word completes a question by the other speaker;
- *nonapplic* for filler phrases such as "in fact", "I mean", etc.

Note that only content words in full sentences were marked for kontrast, and filler phrases such as "in fact" and "I mean" were excluded. A complete description of the annotation guidelines can be found in (Calhoun et al., 2005).

Animacy Each noun and pronoun is labeled for the animacy of its referent (Zaenen et al., 2004). The categories include *concrete*, *non-concrete*, *human*, *org*anizations, *place*, and *time*.

Dialog act Specifies the function of the utterance such as *statement*, *opinion*, *agree*, *reject*, *abandon*; or type of question (*yes/no*, *who*, *rhetoric*)

In addition to the above theoretically motivated features, we used several automatically derivable word measures.

Part-of-speech Two such features were used, the full Penn Treebank tagset (called POS) , and a collapsed tagset (called BroadPOS) with six broad categories (nouns, verbs, function words, pronouns, adjectives and adverbs).

Unigram and bigram probability These features are defined as $log(p_w)$ and $log(p_{w_i}|p_{w_{i-1}})$ respectively and their values were calculated from the Fisher corpus (Cieri et al., 2004). High probability words are less likely to be prominent.

TF.IDF This measure captures how central a word is for a particular conversation. It is a function of the frequency of occurrence of the word in the conversation (n_w), the number of conversations that contain the word in a background corpus (k) and the number of all conversations in the background corpus (N). Formally, *TF.IDF1* $= n_w \times log(\frac{N}{k})$. We

[1] Of all tokens, 8,429 (or 58%) were not accented.

also used a variant, *TF.IDF2*, computed by normalizing *TF.IDF1* by the number of occurrences of the most frequent word in the conversation. *TF.IDF2* = *TF.IDF1*$/max(n_{w \in conv})$. Words with high *TF.IDF* values are important in the conversation and are more likely to be prominent.

Stopword This is a binary feature indicating if the word appears in a high-frequency stopword list from the Bow toolkit (McCallum, 1996). The list spans both function and content word classes, though numerals and some nouns and verbs were removed.

Utterance length The number of words.

Length The number of characters in the words. This feature is correlated with phonetic features that have been shown to be useful for the task, such as the number of vowels or phones in the word.

Position from end/beginning The position of the word in the utterance divided by the number of words that precede the current word.

Accent ratio This final (new) feature takes the "memorization" of previous productions of a given word to the extreme, measuring how likely it is that a word belongs to a prominence class or not. Our feature extends an earlier feature proposed by (Yuan et al., 2005), which was a direct estimate of how likely it is for the word to be accented as observed in some corpus. (Yuan et al., 2005) showed that the original accent ratio feature was not included in the best set of features for accent prediction. We believe the reason for this is the fact that the original accent ratio feature was computed for all words, even words in which the value was indistinguishable from chance (.50). Our new feature incorporates the significance of the prominence probability, assuming a default value of 0.5 for those words for which there is insufficient evidence in the training data. More specifically,

$$AccentRatio(w) = \begin{cases} \frac{k}{n} & \texttt{if } B(k,n,0.5) \leq 0.05 \\ 0.5 & \texttt{otherwise} \end{cases}$$

where k is the number of times word w appeared accented in the corpus, n is the total number of times the word w appeared, and $B(k,n,0.5)$ is the probability (under a binomial distribution) that there are k successes in n trials if the probability of success and failure is equal. Simply put, the accent ratio of a word is equal to the estimated probability of the word being accented if this probability is significantly different from 0.5, and equal to 0.5 otherwise. For example, *AccentRatio(you)=0.3407*, *AccentRatio(education)=0.8666*, and *AccentRatio(probably)=0.5*.

Many of our features for accent prediction are based only on the 12 training conversations. Other features, such as the unigram, bigram, and TF*IDF features, are computed from larger data sources. Accent ratio is also computed over a larger corpus, since the binomial test requires a minimum of six occurrences of a word in the corpus in order to get significance and assign an accent ratio value different from 0.5. We thus used 60 Switchboard conversations (Ostendorf et al., 2001), annotated for pitch accent, to compute k and n for each word.

3 Results

For our experiments we used the J48 decision trees in WEKA (Witten and Frank, 2005). All the results that we report are from 10-fold cross-validation on the 12 Switchboard conversations.

Some previous studies have reported results on prominence prediction in conversational speech with the Switchboard corpus. Unfortunately these studies used different parts of the corpus or different labelings (Gregory and Altun, 2004; Yuan et al., 2005), so our results are not directly comparable. Bearing this difference in mind, the best reported results to our knowledge are those in (Gregory and Altun, 2004), where conditional random fields were used with both textual, acoustic, and oracle boundary features to yield 76.36% accuracy.

Table 1 shows the performance of decision tree classifiers using *a single* feature. The majority class baseline (not accented) has accuracy of 58%. Accent ratio is the most predictive feature: the accent ratio classifier has accuracy of 75.59%, which is two percent net improvement above the previously known best feature (unigram). The accent ratio classifier assigns a "no accent" class to all words with accent ratio lower than 0.38 and "accent" to all other words. In Section 4 we discuss in detail the accent ratio dictionary, but it is worth noting that it does correctly classify even some high-frequency function words like "she", "he", "do" or "up" as accented.

3.1 Combining features

We would expect that a combination of features would lead to better prediction when compared to a classifier based on a single feature. Several past studies have examined classes of features. In order to quantify the utility of different specific features, we ran exhaustive experiments producing classifiers with all possible combinations of two, three, four and five features.

As we can see from figure 1 and table 2, the classifiers using accent ratio as a feature perform best, for all sizes of feature sets. Moreover, the increase of performance compared to a single-feature classifier is very slight when accent ratio is used as feature. Kontrast seems to combine well with accent ratio and all of the best classifiers with more than one feature use kontrast in addition to accent ratio. This indicates that automatic detection of kontrast can potentially help in prominence prediction. But the gains are small, the best classifiers without kontrast but still including accent ratio perform within 0.2 percent of the classifiers that use both.

On the other hand, classifiers that do not use accent ratio perform poorly compared to those that do, and even a classifier using five features (unigram, broad POS, token length, position from beginning and bigram) performs about as well as a classifier using solely accent ratio as a feature. Also, when accent ratio is not used, the overall improvement of the classifier grows faster with the addition of new features. This suggest that accent ratio provides rich information about words beyond that of POS class and general informativeness.[2]

Table 2 gives the specific features in $(n+1)$-feature classifiers that lead to better results than the best n-classifier. The figures are for the classifiers performing best overall. Interestingly, none of these best classifiers for all feature set sizes uses POS or unigram as a feature. We assume that accent ratio captures all the relevant information that is present in the unigram and POS features. The best classifier with five features uses, in addition to accent ratio, kontrast, tf.idf, information status and distance from the beginning of the utterance. All of these features convey somewhat orthogonal information: semantic, topicality, discourse and phrasing information respectively. Still, all of them in combination improve the performance over accent ratio as a single feature only by one percent.

Accent Ratio (AR)	75.59%
AR + Kontrast	76.15%
AR + END/BEG	75.91%
AR + tf.idf2	75.82%
AR + Info Status	75.82%
AR + Length	75.77%
AR + tf.idf1	75.74%
AR + unigram	75.71%
AR + stopword	75.70%
AR + kontrast + length	76.45%
AR + kontrast + BEG	76.24%
AR + kontrast + unigram	76.24%
AR + kontrast + tf.idf1	76.24%
AR + kontrast + length + tfidf1	76.56%
AR + kontrast + length + stopword	76.54%
AR + kontrast + length +tf.idf2	76.52%
AR + kontrast + Status + BEG	76.47%
AR + kontrast + tf.idf1 + Status + BEG	76.65%
AR + kontrast + tf.idf2 + Status + BEG	76.58%

Table 2: Performance increase augmenting the accent ratio classifier.

Figure 1 shows the overall improvement of classifiers with the addition of new features in three scenarios: overall best, best when kontrast is not used as a feature and best with neither kontrast nor accent ratio. The best classifier with five features that do not include kontrast has accent ratio, broad POS, word length, stopword and bigram as features and has accuracy of 76.28%, or just 0.27% worse than the overall best classifier that uses kontrast and information status. This indicates that while there is some benefit to using the two features, they do not lead to any substantial boost in performance. Strikingly, the best classifier that uses neither accent ratio nor kontrast performs very similarly to a classifier using accent ratio as the only feature: 75.82% for the classifier using unigram, POS, tf.idf1, word length and position from end of the utterance.

3.2 The power of linguistic features

One of the objectives of our study was to assess how useful gold-standard annotations for complex linguistic features are for the task of prominence prediction. The results in this section indicate that animacy distinctions (concrete/non-concrete, person, time, etc) and dialog act did not have much power

[2] To verify this we will examine the accent ratio dictionary in closer detail in the next section.

AccentRatio	**unigram**	**stopword**	**POS**	**tf.idf2**	**tf.idf1**	**BroadPos**	**Length**	**Kontrast**	**bigram**	**Info Stat**
75.59	73.77	70.77	70.28	70.14	69.50	68.64	67.64	67.57	65.87	64.13

Table 1: Single feature classifier performance. Features not in the table (position from end, animacy, utterance length and dialog act) all achieve lower accuracy of around 60%

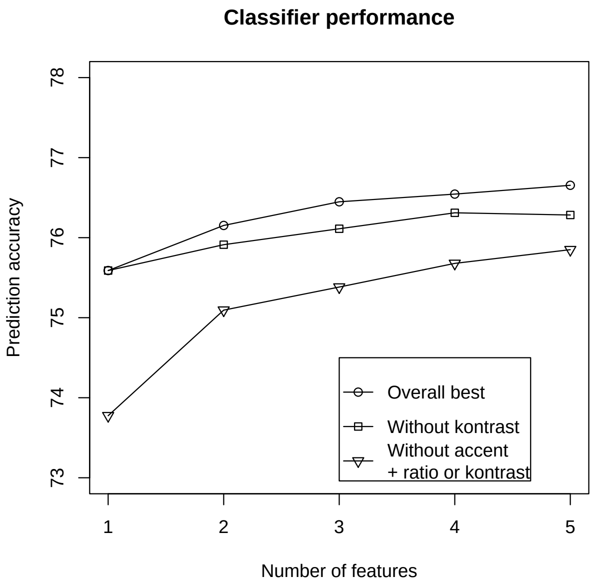

Figure 1: Performance increase with the addition of new features.

as individual features (table 1) and were never included in a model that was best for a given feature set size (table 2).

Information status is somewhat useful and appears in the overall best classifier with five features (table 2). But when compared with other classifiers with the same number of features, the benefits from adding information status to the model are small. For example, the accent ratio + information status classifier performs 0.23% better than accent ratio alone, but so does the classifier using accent ratio and tf.idf. There are two reasons that can explain why the givenness of the referent is not as helpful as we might have hoped. First of all, the information status distinction applies only to referring expressions and has undefined values for words such as verbs, adjectives or function words. Second, information status of an entity influences the form of referring expression that is used, with *old* items being more likely to be pronominalized. In the numerous cases where pronominalization of *old* information does occur, features such as POS, unigram or accent ratio will be sensitive to the change of information status simply based on the lexical item.

Kontrast is by far the most useful linguistic feature. It is used in all of the best classifiers for any feature set size (table 2). It applies to more words than givenness does, since salience distinctions can be made for any part-of-speech class. Still, not all words were annotated for kontrast either, and moreover kontrast only captures one kind of semantic salience. This is particularly true of discourse markers like “especially” or “definitely”: these would either be in sentence fragments that weren’t marked for kontrast, or would probably be marked as ’background’ since they are not salience triggers in a semantic sense. As we can see from figure 1, classifiers that use kontrast perform only slightly better than others that use only “cheaper” features.

4 The accent ratio dictionary

Contrary to our initial expectations, both classes in the accent ratio dictionary (for both low and high probability of being prominent) cover the full set of possible POS categories. Tables 3 and 4 list words in both classes (with words sorted by increasing accent ratio in each column). The “no accent” class is dominated by function words, but also includes nouns and verbs. One of the drawbacks of POS as a feature for prominence prediction is that normally auxiliary verbs will be tagged as “VB”, the same class as other more contentful verbs. The informativeness (unigram probability) of a word would distinguish between these types of verbs, but so does the accent ratio measure as well.

Furthermore, some relatively frequent words such as “too”, “now”, “both”, “no”, “yes”, “else”, “wow” have high accent ratio, that is, a high probability for accenting. Such distinctions within the class of function words would not be possible on the basis of in-

.00–.08	.09–.16	.17–.24	.25–.32	.33–.42
a	could	you’d	being	me
uh	in	because	take	i’ve
um	minutes	oh	said	we’re
uh-huh	and	since	wanna	went
the	by	says	been	over
an	who	us	those	you
of	grew	where	into	thing
to	cause	they’ve	little	what
were	gonna	am	until	some
as	about	sort	they’re	out
than	their	you’re	I	had
with	but	didn’t	that	make
at	on	her	don’t	way
for	be	going	this	did
from	through	i’ll	should	anything
or	which	will	type	i’m
you’ve	are	our	we	kind
was	we’ll	just	so	go
would	during	though	have	stuff
it	huh	like	got	then
when	is	your	new	she
them	bit	needs	mean	he
it’s	there’s	my	much	do
if	any	many	i’d	up
can	has	they	know	
him	stayed	get	doesn’t	
these	supposed	there	even	

Table 3: Accent ratio entries with low prominence probability.

formativeness, POS, or even information structure features. Another class like that is words like “yes”, “okay”, “sure” that are mostly accented by virtue of being the only word in the phrase.

Some rather common words, “not” for example, are not included in the accent ratio dictionary because they do not exhibit a statistically strong preference for a prominence class. The accent ratio classifier would thus assign class “accented” to the word “not”, which is indeed the class this word occurs in more often.

Another fact that becomes apparent with the inspection of the accent ratio dictionary is that while certain words have a statistically significant preference for deaccenting, there is also a lot of variation in their observed realization. For example, personal pronouns such as “I” and “you” have accent ratios near 0.33. This means that every third such pronoun was actually realized as *prominent* by the speaker. In a conversational setting there is an implicit contrast between the two speakers, which could partly explain the phenomenon, but the situations which prompt the speaker to realize the distinction in their speech will be the focus of a future linguistic investigation.

.58–.74	.75–.79	.80–.86	.87–1.0
lot	both	sometimes	half
time	no	change	topic
now	seems	child	else
kids	life	young	obviously
old	tell	Texas	themselves
too	ready	town	wow
really	easy	room	gosh
three	heard	pay	anyway
work	isn’t	interesting	Dallas
nice	again	true	outside
yeah	first	mother	mostly
two	right	problems	yes
person	children	agree	great
day	married	war	exactly
working	may	needed	especially
job	happen	told	definitely
talking	business	finally	lately
usually	still	neat	thirty
rather	daughter	sure	higher
places	gone	house	forty
government	guess	okay	hey
ten	news	seven	Iowa
parents	major	best	poor
paper	fact	also	glad
actually	five	older	basic

Table 4: Accent ratio values for words with high probability for being accented.

Kontrast is helpful in predicting “accented” class for some generally low ratio words. However, even with its help, production variation in the conversations cannot be fully explained. The following examples from our corpus show low accent ratio words (that, did, and, have, had) that were produced as prominent.

so i did THAT. and then i, you know, i DID that for SIX years. AND then i stayed HOME with my SON.

i HAVE NOT, to be honest, HAD much EXPERIENCE with CHILDREN in that SITUATION.

they're going to HAVE to WORK it OUT to WORKING part TIME.

The examples attest to the presence of variation in production: in the first utterance, for example, we see the words “did”, “and” and “that” produced both as prominent and not prominent. Intonational phrasing most probably accounts for some of this variation since it is likely that even words that are typically not prominent will be accented if they occur just before or after a longer pause. We come back to this point in the closing section.

5 Robustness of accent ratio

While accent ratio works well for our data (Table 2), a feature based so strongly on memorizing the status of each word in the training data might lead to problems. One potential problem, suggested by Pan et al. (2002) for lexicalized features in general, is whether a lexical feature like accent ratio might be less robust across genres. Another question is whether our definition of accent ratio is better than one that does not use the binomial test: we need to investigate whether these statistical tests indeed improve performance. We focus on these two issues in the next two subsections.

Binomial test cut-off

As discussed above, the original accent ratio feature (Yuan et al., 2005) was based directly on the fraction of accented occurrences in the training set. We might expect such a use of raw frequencies to be problematic. Given what we know about word distributions in text (Baayen, 2001), we would expect about half of the words in a big corpus to appear only once. In an accent ratio dictionary without binomial test cut-off, all such words will have accent ratio of either exactly 1 or 0, but one or even few occurrences of a word would not be enough to determine statistical significance. By contrast, our modified accent ratio feature uses binomial test cut-off to make the accent ratio more robust to small training sets.

To test if the binomial test cut-off really improved the accent ratio feature, we compared the performance on Switchboard of classifiers using accent ratio with and without cut-off. The binominal test improved the performance of the accent ratio feature from 73.49% (Yuan *et al.* original version) to 75.59% (our version).

Moreover, Yuan *et al.* report that their version of the feature did not combine well with other features, while in our experiments best performance was always achieved by the classifiers that made use of the accent ratio feature in addition to others.

A cross-genre experiment: broadcast news

In a systematic analysis of the usefulness of different informativeness, syntactic and semantic features for prominence prediction, Pan et al. (2002) showed that *word identity* is a powerful feature. But they hypothesized that this would not be a useful feature in a domain independent pitch accent prediction task. Their hypothesis that word identity cannot be a robust across genres would obviously carry over to accent ratio. In order to test the hypothesis, we used the accent ratio dictionary derived from the Switchboard corpus to predict prominence in the Boston University Radio corpus of broadcast news. Using an accent ratio dictionary from Switchboard and assigning class “not accented” to words with accent ratio less than 0.38 and “accented” otherwise leads to 82% accuracy of prediction for this broadcast news corpus. If the accent ratio dictionary is built from the BU corpus itself, the performance is 83.67%.[3] These results indicate that accent ratio is a robust enough feature and is applicable across genres.

6 Conclusions and future work

In this paper we introduced a new feature for prominence prediction, accent ratio. The accent ratio of a word is the (maximum likelihood estimate) probability that a word is accented if there is a significant preference for a class, and 0.5 otherwise. Our experiments demonstrate that the feature is powerful both by itself and in combination with other features. Moreover, the feature is robust to genre, and accent ratio dictionaries can be used for prediction of prominence in read news with very good results.

Of the linguistic features we examined, kontrast is the only one that is helpful beyond what can be gained using shallow features such as n-gram probability, POS or tf.idf. While the improvements from kontrast are relatively small, the consistency of these small improvements suggest that developing automatic methods for approximating the gold-standard annotation we used here, similar to what has been done for information status in (Nissim, 2006), may be worthwhile. An automatic predictor for kontrast may also be helpful in other applications such as question answering or textual entailment.

All of the features in our study were text-based. There is a wide variety of research investigating phonological or acoustic features as well. For example Gregory and Altun (2004) used acoustic features

[3] This result is comparable with the result of (Yuan et al., 2005) who in their experiment with the same corpus report the best result as 83.9% using *three* features: unigram, bigram and backwards bigram probability.

such as duration and energy, and phonological features such as oracle (hand-labeled) intonation phrase boundaries, and the number of phones and syllables in a word. Although acoustic features are not available in a text-to-speech scenario, we hypothesize that in a task where such features are available (such as in speech recognition applications), acoustic or phonological features could improve the performance of our text-only features. To test this hypothesis, we augmented our best 5-feature classifier which did not include kontrast with hand-labeled intonation phrase boundary information. The resulting classifier reached an accuracy of 77.45%, more than one percent net improvement over 76.28% accuracy of the model based solely on text features and not including kontrast. Thus in future work we plan to incorporate more acoustic and phonological features.

Finally, prominence prediction classifiers need to be incorporated in a speech synthesis system and their performance should be gauged via listening experiments that test whether the incorporation of prominence leads to improvement in synthesis.

References

R. H. Baayen. 2001. *Word Frequency Distributions*. Kluwer Academic Publishers.

D.L. Bolinger. 1961. Contrastive Accent and Contrastive Stress. *Language*, 37(1):83–96.

J. Brenier, A. Nenkova, A. Kothari, L. Whitton, D. Beaver, and D. Jurafsky. 2006. The (non)utility of linguistic features for predicting prominence in spontaneous speech. In *IEEE/ACL 2006 Workshop on Spoken Language Technology*.

G. Brown. 1983. Prosodic structure and the given/new distinction. *Prosody: Models and Measurements*, pages 67–77.

S. Calhoun, M. Nissim, M. Steedman, and J.M. Brenier. 2005. A framework for annotating information structure in discourse. *Pie in the Sky: Proceedings of the workshop, ACL*, pages 45–52.

W. Chafe. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and Topic*, pages 25–55.

C. Cieri, D. Graff, O. Kimball, D. Miller, and Kevin Walker. 2004. Fisher English training speech part 1 transcripts. *LDC*.

J. Godfrey, E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *IEEE ICASSP-92*.

M. Gregory and Y. Altun. 2004. Using conditional random fields to predict pitch accents in conversational speech. *Proceedings of ACL*, 2004.

J. Gundel, N. Hedberg, and R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69:274–307.

J. Hirschberg. 1993. Pitch Accent in Context: Predicting Intonational Prominence from Text. *Artificial Intelligence*, 63(1-2):305–340.

A. McCallum. 1996. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. http://www.cs.cmu.edu/ mccallum/bow.

M. Nissim, S. Dingare, J. Carletta, and M. Steedman. 2004. An annotation scheme for information status in dialogue. In *LREC 2004*.

M. Nissim. 2006. Learning information status of discourse entities. In *Proceedings of EMNLP 2006*.

M. Ostendorf, I. Shafran, S. Shattuck-Hufnagel, L. Carmichael, and W. Byrne. 2001. A prosodically labeled database of spontaneous speech. *Proc. of the ISCA Workshop on Prosody in Speech Recognition and Understanding*, pages 119–121.

S. Pan and J. Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of ACL-00*.

S. Pan and K. McKeown. 1999. Word informativeness and automatic pitch accent modeling. In *Proceedings of EMNLP/VLC-99*.

S. Pan, K. McKeown, and J. Hirschberg. 2002. Exploring features from natural language generation in prosody modeling. *Computer speech and language*, 16:457–490.

E. Prince. 1992. The ZPG letter: subject, definiteness, and information status. In S. Thompson and W. Mann, editors, *Discourse description: diverse analyses of a fund raising text*, pages 295–325. John Benjamins.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116.

E. Vallduví and M. Vilkuna. 1998. On rheme and kontrast. *Syntax and Semantics*, 29:79–108.

I. H. Witten and E. Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. 2nd Edition, Morgan Kaufmann, San Francisco.

J. Yuan, J. Brenier, and D. Jurafsky. 2005. Pitch Accent Prediction: Effects of Genre and Speaker. *Proceedings of Interspeech*.

A. Zaenen, J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M.C. O'Connor, and T. Wasow. 2004. Animacy Encoding in English: why and how. *ACL Workshop on Discourse Annotation*.