FORMS OF ENGLISH FUNCTION WORDS—EFFECTS OF DISFLUENCIES, TURN POSITION, AGE AND SEX, AND PREDICTABILITY

Alan Bell*, Daniel Jurafsky*, Eric Fosler–Lussier†, Cynthia Girand*, and Daniel Gildea†
*Department of Linguistics, University of Colorado, Boulder
†International Computer Science Institute and Computer Science Division, University of California, Berkeley

ABSTRACT

This study examines the role of several non–phonetic factors in the reduction of ten frequent English function words (*I, and, the, that, a, you, to, of, it,* and *in*) in the phonetically-transcribed portion of the Switchboard corpus of spontaneous telephone conversations. Using ordinary linear and logistic regression models, we examined the length of the words and whether their vowels were full or reduced. We show that function words are more likely to be longer or unreduced when they are turn–initial or utterance–final, when the speaker is female (mostly but not completely due to slower rate of speech) and when the word is surprising given the previous or following words. Finally, focusing on finer details of the effect of planning problems on reduction, we show that filled pauses (*uh* and *um*) are the strongest factor in predicting lengthening of a previous function word. The results bear on issues in speech recognition and models of speech production.

1. INTRODUCTION

This study reports on our continuing investigation into a number of non–phonetic factors affecting the reduction of ten of the most frequent English words—*I, and, the, that, a, you, to, of, it,* and *in*—in the Switchboard corpus of conversational telephone speech. Frequent function words are of particular interest because they are not only subject to the contextual and stylistic processes that govern the variation of content word forms, but also typically exhibit additional variation, especially a greater propensity toward reduced forms. Keating (1997) showed that this greater variation in function words was even more evident in Switchboard than in read–speech databases like TIMIT.

Our study sheds more light on the role of two factors that we looked at in Jurafsky et al. (1998):

- **planning problems:** whether the speaker was having difficulty in production, as previously indicated by repetitions, pauses, and use of *um* and *uh*. In this paper we studied the differential effect of these indicators.
- **predictability:** the predictability of the function word, modeled by its probability given the previous two words. In this paper we studied the effect of the following word.

We also report on two new factors that play a role in the reduction of function words.

- **age and sex:** the age and sex of the speaker
- **position in turn and utterance:** whether the utterance was initial or final in the utterance and the turn.

2. METHODOLOGY

2.1 Data

The Switchboard corpus of telephone conversations between strangers was collected in the early 1990's (Godfrey et al. 1992). The corpus contains 2430 conversations averaging 6 minutes each, totaling 240 hours of speech and 3 million words. Approximately four hours of this speech was phonetically hand–transcribed by Greenberg et al. (1996). The speech files were automatically segmented into pseudo–utterances at turn boundaries or at silences of 500 ms or more. The transcribers were given these utterances, the word transcription, and a rough automatic phonetic transcription. They then corrected this rough phonetic transcription, using an augmented version of the arpabet. In general we relied on these Berkeley transcriptions for our coding, although we did listen to, recode, and/or eliminate certain observations;see Jurafsky et al. (1998) for more details on the data coding.

2.2 Dependent Variables

We examined two dependent factors reflecting reduction: **vowel quality** and **word length**. Vowel quality was a categorical variable, coding whether a vowel was **full** or **reduced**. The reduced vowels were ə (arpabet [ax]), ɨ (arpabet [ix]), ɚ (arpabet [axr]), and ɵ (not in the arpabet). Word length was the duration of the word in milliseconds.

2.3 Regression models

We used regression models to evaluate the effects of these factors on the measures of reduction, logistic regression for the categorical variable of vowel quality, ordinary linear regression for length. Thus when we report that an effect was significant, it is meant to be understood that it is a significant parameter in a model that also includes the other significant variables. In other words, after accounting for the effects of the other variables, adding the variable in question produced a significantly better account of the variation. Our models were based on roughly 7300 to 8400 observations.

Logistic regression models the effect of explanatory variables on a categorical variable in terms of the **odds** of the category, which is the ratio of P(category) to 1–P(category). For a binary category like full versus reduced vowel,we estimate the odds by the ratio of the percentages of the two values: the article *a* occurs with a full vowel 24% of the time, and with a reduced vowel 76%; the odds of a full vowel are 24/76 = 0.3 (to one).

3. OUR PREVIOUS RESULTS

In our previous work, we found four factors which significantly affected reduction. We summarize those findings here. Note that our previous work investigated a wider range of dependent variables, including whether the **coda–consonant** was present in *it*, *that*, *of*, and *and*, and also distinguished two subtypes of unreduced pronunciations: 'basic' or 'canonical' pronunciations ([ænd] for *and*) and 'other full' pronunciations ([ɛnd] for *and*).

3.1 Rate of Speech

Speech researchers have long noted the association between faster speech, informal styles, and more reduced forms. We measured rate of speech at a given function word by taking the number of syllables per second in the pause–bounded region immediately surrounding the word; see Fosler–Lussier and Morgan (1998) for more details on this coding.

Unsurprisingly, rate of speech affected all measures of reduction. Comparing the difference between a relatively fast rate of 7.5 syllables per second and a slow rate of 2.5 syllables per second, the estimated increase in the odds of full to reduced vowels was 3.7, i.e. the odds of a full vowel at the slow rate was 3.7 times the odds at the faster rate ($p < .0001$).

3.2 Planning Problems

The production of speech is accompanied by a variety of disfluencies. Some of these disfluencies are prospective, largely due to speakers' trouble in formulating an idea, and expressing it with the proper syntax, words, prosody, and articulation. Fox Tree and Clark (1997) suggested that such 'planning problems' cause words in immediately preceding speech to have less reduced pronunciations. They found this to be true for *the*.

We investigated the effects of planning problems on the ten function words. Following earlier research, we took **disfluencies** (pauses, filled pauses like *uh* or *um*, and repetitions) to be symptoms of planning problems; any function word followed by one of these was assumed to belong to a planning problem context. Disfluencies were not infrequent, although they did vary by function word, as Table 1 shows.

a	*the*	*to*	*in*	*of*	*and*	*that*	*I*	*it*	*you*
8.7	11.7	7.1	7.8	7.7	22.6	19.0	11.0	12.9	3.5

Table 1: Percentage of functors occurring before disfluencies.

The effect of a planning problem on word length was massive and across–the–board (see Figure 1). The effect, both overall and for each word, remains after partialing out effects of rate, predictability, and next consonant/vowel ($p < .0001$). Words are roughly twice as long before a disfluency than before a word. All classes of vowels, basic, full, and reduced, are lengthened.

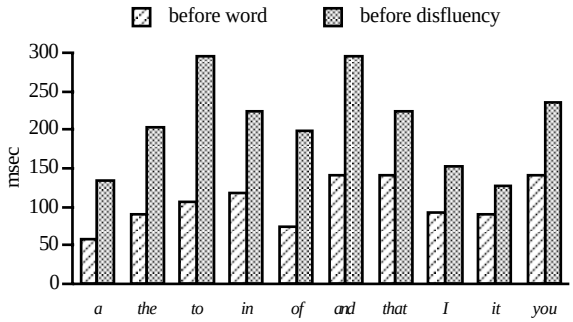

Figure 1. Average length for function words when followed by another word or by a disfluency

3.3 Following Consonant/Vowel

A general fact about weakening processes is that the form of a word is influenced by the segmental context. More reduced forms tend to occur before a consonant than before a vowel. This is often modeled as allomorphy, e.g. *the* as [ði] before vowels but [ðə] before consonants (Keating 1994).

We found significantly less reduction (in all four variables) when the next word began with a vowel than when it began with a consonant. As expected, the odds of a basic [ði] form of *the* were greatly increased before a vowel. However, *to* and *of* were also similarly affected,suggesting an allomorphic account of a [tu]/[tə] or [ʌv]/[əv] alternation.

3.4 Predictability and Collocation

Jespersen (1923) notes that the predictability of the word in its context is an important factor contributing to weakened pronunciations. To measure predictability, we estimated the log of the conditional probability of a function word given the previous two words using a backoff trigram grammar with Good–Turing discounting trained over the entire Switchboard corpus. Greater predictability increases the likelihood of reduction. More predictable words were shorter than less predictable words ($p < .0001$).

We also found collocational effects; the *you* of *you know* was highly reduced, and the *of* in partitive constructions (*kind of*, *lots of*, etc.) was significantly more likely ($p < .001$) to have no coda than in other uses (such as *thought of*, *outside of*). This suggests that the partitive construction may be stored or unitized as a mental routine.

4. ACCENT

One of the strongest factors influencing the pronunciation of a word is whether it receives accent. Since function words are relatively unlikely to be accented, the lack of prosodic coding in our corpus is not a great handicap. Nevertheless, when accent does fall on a functor, it will be longer and have a fuller pronunciation, and could thus be responsible for some of the above effects. We listened to 120 instances of the functors, concentrating on their longest full vowel tokens. Each item was coded for presence or absence of pitch accent. About

one–third of the items were coded independently by two listeners, and of the other two–thirds, any low–confidence tokens were considered by the other listener also.

Only four functors had more than one accented token out of ten: *I*, 12 of 20; *you*, 7 of 15; *that*, 4 of 15; and *and*, 2 of 10. Since these are sampled from the longer tokens, overall frequencies of accent will be much lower. For these items we cannot rule out the possibility that accented tokens occur disproportionately in one or more of our predictor contexts. For the disfluency contexts, accented *and* and *that* occurred equally in our sample before words or before disfluencies; accented *I* and *you* occurred slightly more frequently in the disfluent contexts, but the difference for this small sample was far from significant. We cannot rule out that accent may be found to interact with our other effects in a larger sample, of course, but the magnitude and the scope of its influence on our results appears to be limited.

5. AGE AND SEX

Previous research has shown that sex is an important factor in pronunciation variation. Byrd (1994), in her study of read speech in the TIMIT corpus, found that men spoke on average 6.2% faster than women. Shriberg (1999), in her study of disfluencies in Switchboard, found that men had slightly more disfluencies per word than women. Besides this effect of sex, we were interested in the role of age in pronunciation variation. The age of Switchboard speakers at time of recording ranged from 17 to 68.

Because of these previous results, we suspected that the effects of sex and age on pronunciation would mostly be realized through changes in speaking rate and disfluency rate. We thus began by investigating the role of sex and age on speaking rate and on the incidence of disfluencies. We then looked at their remaining marginal effects on pronunciation.

5.1 Effect of age and sex on rate.

There is a large effect of sex on speaking rate ($p < .0001$). On average men spoke 6.4% faster than women. Men had an average rate of 5.35 syllables per second, and women an average rate of 5.03 syllables per second. It is somewhat surprising that such a similar difference is found for both read speech and conversation.

We also found an effect of age on rate ($p < .0001$). Older speakers spoke more slowly; 5.12 syllables per second for speakers over 50, compared to 5.44 syllables per second for speakers under 30.

Finally, there was an interaction of age and sex. While women on average spoke more slowly than men, older women spoke even more slowly than older men. Figure 2 shows all these effects of age and sex on rate.

5.2 Effect of age and sex on disfluencies.

As expected from Shriberg's (1999) research on the Switchboard corpus, we found an effect of sex on disfluencies. Men had a 16% higher odds of disfluencies than women. (The raw rate of disfluency was 12.2% per word for men, and 10.9% for women.) Although there was no main effect of age, age and sex again interacted. While younger men had a greater rate of disfluencies than women (the odds of disfluencies were 64% greater for 20–year old men than 20–year–old women), the difference decreased with age, so that older men had approximately the same disfluency rate as women.

If the effects of sex and age on speaking rate and disfluencies are partialed out, is there any remaining effect on pronunciation? There was a remaining small effect of sex ($p = .030$) on word length (men's words were still 3% shorter), but no remaining effect of age. For vowel reduction, however, the remaining effect of sex is much stronger ($p < .0001$).

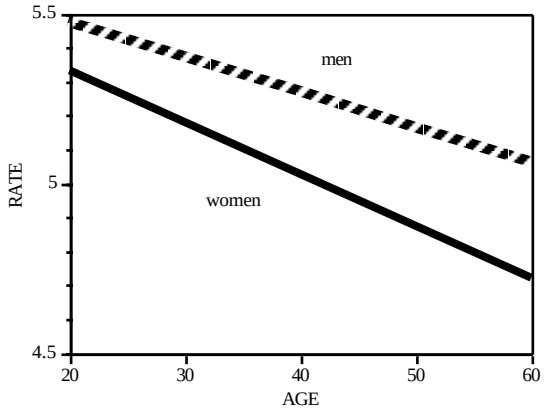

Figure 2. Estimated rate of speech in syllables per second for men and women by age.

6. PREDICTABILITY

Our previous work showed that words are shorter and more reduced when they are predictable (i.e. when they have a high trigram probability given the previous two words). We were interested in whether the *following* word could play a role in predictability, since the speaker is presumably already planning the next word when the current word is articulated.

We examined the **centered trigram** probability (the log probability of a function word given the previous and the following word using the methods described in §3.4. We found significant effects for both the trigram and centered trigram probabilities. Function words averaged 83 ms shorter when they were highly predictable by a standard trigram ($p(word) = .56$) than when they were highly unpredictable by a standard trigram ($p(word) = .0056$). After partialing out this effect of standard trigram, function words were still 31 ms shorter when they were highly predictable by a centered trigram ($p(word) = .56$) than when they were highly unpredictable by a centered trigram ($p(word) = .00000056$).

7. POSITION IN TURN AND UTTERANCE

Previous research has shown the importance of the position of a word in its turn or utterance, particularly of initial and final position. Phrase–final lengthening is a well–known factor. Research on repair (Fox and Jasperson 1995 inter alia) has suggested that planning problems may tend to be located early in turns. An analysis of errors in automatic speech recognition

of Switchboard (Jurafsky et al. 1998b) showed that utterance–initial words were more often misrecognized.

We examined the effect of word position on reduction and on the likelihood of disfluencies. We coded each word for its position in the turn and in the utterance, following the definitions of turn and utterance used in the Penn Treebank annotation of Switchboard (Meteer et al 1995).

Both turn–initial and utterance–final words are significantly longer than other words. Much of the effect of turn position on length is realized via an increased disfluency rate (disfluency odds ratio of 1.41 for turn–initial position, and of 4.42 for utterance–final position). (Unlike Shriberg (1999), we did not find that disfluencies are significantly more likely to occur utterance–initially.) But even after factoring out the effect of disfluencies on reduction, turn–initial words were still 20% longer than non–turn–initial words, and utterance–final words were still 35% longer than non–utterance–final words.

8. DISFLUENCIES

Does the powerful effect of disfluent items on the preceding function word extend to each of the categories—pause, filled pause, and repetition—that we used as indicators of planning problems? Earlier work by O'Shaughnessy (1992) and Girand et al. (1998) suggested that it might not, since they found that much of the lengthening of a word preceding its repetition was due to lengthening occurring when pauses fell between the repeated words.

An item analysis for the three categories found somewhat different results. First, silent pauses, filled pauses, and repetition each contribute strongly to the longer and less reduced forms found before them. Second, filled pauses, not silent pauses, have the strongest effect. Third, while the effect of silence was greater than that of repetition, it was only marginally significant ($p = .040$). Although our assessment of these effects was carried out after partialing out effects of rate,

	following item			
	word	pause	filled pause	repetition
word length	113	212	353	183
full vowel odds	1.33	3.66	12.20	6.99

Table 2. Average length and odds of a full vowel for functors preceding words and disfluency contexts.

for simplicity we summarize them by reporting raw durations and odds of reduction in the table below. The relatively smaller effect for pauses is not necessarily inconsistent with the prior results; since they concerned pauses *within* repetition strings.

9. CONCLUSION

This study has measured the affect of a number of non–phonetic factors on the reduction of ten English function words in the Switchboard corpus. Function words are more likely to be unreduced when they are turn–initial or utterance–final. Female speakers are more likely to have long or unreduced function words; this effect is mostly but not completely due to their slightly slower speaking rate. Finally, focusing on finer details of the effect of planning problems on reduction, we show that filled pauses (*uh* and *um*) are the strongest factor in predicting lengthening of a previous function word.

It is interesting that these factors are all extremely local, mostly involving the immediately previous or following word, an immediately previous or following turn or utterance boundary, or the immediate rate of speech in a very local region. In terms of cognitive models of speech production, while some of these effects are likely to arise at the levels of lexical retrieval and compilation of syntactic and prosodic frames, others will require more immediate access to the articulatory routines. Whatever the sources of the effects, the results hold out the hope for speech engineering of building good predictive models of word pronunciation based only on very local information.

ACKNOWLEDGEMENTS

This project was partially supported by NSF IIS-9733067 and the Center for Language and Speech Processing at The Johns Hopkins University. Many thanks to Elizabeth Shriberg and Steve Greenberg.

REFERENCES

[1] Byrd, Dani. 1994. Relations of sex and dialect to reduction. S*peech Communication* 15, 39–54.
[2] H. H. Clark and T. Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology*.
[3] E. Fosler–Lussier and N. Morgan. 1998. Effects of speaking rate and word frequency on conversational pronunciations. Proceedings of the ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition, Kekrade (Netherlands).
[4] B. Fox and R. Jasperson. 1995. A syntactic exploration of repair in English conversation. In P. Davis, editor, *Descriptive and Theoretical Modes in the Alternative Linguistics*, 77–134. Benjamins, Amsterdam.
[5] J. E. Fox Tree and H. H. Clark. 1997. *Pronouncing* ''the'' as ''thee'' to signal problems in speaking. Cognition, 62:151–167.
[6] C. Girand, A. Bell, D. Jurafsky, and E. Fosler–Lussier. 1998. The structure of repetition strings in the Switchboard corpus. *JASA*, 104:3, 1819.
[7] J. Godfrey, E. Holliman, and J. McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. *ICASSP–92*, 517–520.
[8] S. Greenberg, D. Ellis, and J. Hollenback. 1996. Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *ICSLP–96*, Philadelphia.
[9] S. Greenberg. 1998. Speaking in shorthand – A syllable–centric perspective for understanding pronunciation variation. Proceedings of the ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition, Kekrade, Netherlands, 47–56.
[10] O. Jespersen. 1923. *Language*. Henry Holt, New York.
[11] D. Jurafsky, A. Bell, E. Fosler–Lussier, C. Girand, and W. Raymond. 1998. Reduction of English function words in Switchboard. In *ICSLP–98* volume 7, 3111–3114.
[12] D. Jurafsky, R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E Shriberg, A Stolcke, P Taylor, C. Van Ess–Dykema. 1998b. Switchboard Discourse Language Modeling Report. Research Note 30, Center for Speech and Language Processing, Johns Hopkins University.
[13] P. A. Keating. 1997. Word–level phonetic variation in large speech corpora. Ms, Berlin Conference on the 'Phonetic Word'.
[14] P. A. Keating, D. Byrd, E. Flemming, and Y. Todaka. 1994. Phonetic analysis of word and segment variation using the timit corpus of American English. *Speech Communication*, 14:131–142.
[15] M. Meteer et al. 1995. Disfluency Annotation Stylebook for the Switchboard Corpus. Linguistic Data Consortium.
[16] D. O'Shaughnessy. 1992. Recognition of hesitations in spontaneous speech. *Proc. IEEE ASSP Conf.*, 521–524.
[17] E. Shriberg. 1999 to appear. *Disfluencies: Computational and Statistical Models of Spontaneous Speech*. John Benjamins, Amsterdam.