5th International Conference on Spoken
Language Processing (ICSLP 98)
Sydney, Australia
November 30 - December 4, 1998

REDUCTION OF ENGLISH FUNCTION WORDS IN SWITCHBOARD

Daniel Jurafsky, Alan Bell*, Eric Fosler-Lussier†, Cynthia Girand*, and William Raymond**

University of Colorado, Boulder*
University of California, Berkeley†

ABSTRACT

The causes of pronunciation reduction in 8458 occurrences of ten frequent English function words in a four-hour sample from conversations from the Switchboard corpus were examined. Using ordinary linear and logistic regression models, we examined the length of the words, the form of their vowel (basic, full, or reduced), and final obstruent deletion. For all of these we found strong, independent effects of speaking rate, predictability, the form of the following word, and planning problem disfluencies. The results bear on issues in speech recognition, models of speech production, and conversational analysis.

1. INTRODUCTION

This study reports the results of an investigation of some factors affecting the reduction or lenition of ten of the most frequent English words, namely *I, and, the, that, a, you, to, of, it,* and *in*, in the Switchboard corpus of conversational 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. We mainly report on the effects of four factors that, based on earlier research, might affect reduction:

rate of speech: the rate of speech of the current utterance in syllables/second.

planning problems: whether the speaker was having difficulty in production, as indicated by repetitions, pauses, and use of the fillers *um* and *uh*.

segmental context: whether the onset of the following word was a vowel or consonant.

predictability: the predictability of the function word in its context, as modeled by its conditional probability given the previous two words.

Preliminary remarks on the effects of other factors such as position in utterance, following syllable, and collocational effects are included; other potentially important factors such as syntactic function and dialect are not addressed.

2. METHODOLOGY

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.

Because we were particularly interested in the difference between full and non-full forms, we examined three dependent factors reflecting various processes of lenition, reduction, or shortening: (Throughout this paper we will use the term 'reduced' to refer to these more elliptical forms.)

vowel quality: We coded each vowel as **basic**, **other full**, or **reduced**. The basic vowel was the citation or clarification pronunciation, e.g. [ði] for *the*. The reduced vowels were [ə] (arpabet [ax]), [ɨ] (arpabet [ix]), [ɚ] (arpabet [axr]), and [ɵ] (not in the arpabet).[1] Any other vowel was a full vowel. This three-way distinction was split into two binary contrast variables: full/reduced (basic and other full vowel versus reduced vowel) and basic/full.

coda consonant: for words which have coda obstruents (*it, that, and, of*), whether the consonant is deleted. (The sonorant nasal codas of *in* and *and* were not considered.)

length: the duration of the word in milliseconds.

We used regression models to evaluate the effects of these factors on the measures of reduction, logistic regression for the categorical variables of vowel quality and coda presence, 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.[2]

Logistic regression models the effect of explanatory variables on a categorical variable in terms of the **odds** of the category, which is the ratio $\frac{P(\text{category})}{1-P(\text{category})}$. For a binary category like full

---

[1] In general we relied on Berkeley transcriptions for our coding. We did listen to the utterances in five classes of tokens that seemed likely to affect our analysis: possible misalignments in our processing, a sample of tokens transcribed as having no segment, all tokens of arpabet [ux], all tokens of arpabet [er], and a random sample of 100 of the function words. Some items were recoded, mainly [ux] as either a non-reduced high front round vowel [ʉ], as prescribed, or reduced [ɵ], or [er] as either full [ɝ] or reduced [ɚ]. Some items were removed, mainly those transcribed as having no segment, since from our sample we judged that many were equally segmental as other transcriptions. Our judgements of the tokens in the random sample in general agreed with the original transcribers. Notably, however, we judged five of the 57 full vowels in the sample to be reduced, whereas we agreed with the coding of all the reduced vowels. This suggests that there may be a bias toward full vowels in the transcription.

[2] The number of items for regression analyses was always less than the total of 8458 items, because for each variable we eliminated some problematic and extreme values. The regressions for length and full/reduced were based on 7791 observations; for basic/full (reduced vowels omitted), 4695 observations; and for coda (over four words), 2763 observations.

	Basic	**Other Full**	**Reduced**
a	[eɪ]	[ʌ],[ɪ]	[ə],[ɨ]
the	[ði],[i],[di]	[ðʌ],[ðɪ],[ʌ]	[ðə],[ðɨ], [ə]
in	[ɪn],[ɪ],[ɪɾ̃],	[ɛn],[ʌn],[æn]	[ɨn],[n̩],[ən]
of	[ʌv],[ʌ],[ʌvv]	[ɪ],[i],[ɑ]	[ə],[əv],[əf]
to	[tu],[tʉ],[ɾu]	[tʊ],[tɪ],[tʌ]	[tə],[tɨ],[ə]
and	[æn],[ænd],[æɾ̃]	[ɛn],[ɪn],[ʌn]	[ɨn],[n̩],[ən]
that	[ðæ],[ðæt], [æ]	[ðɛ],[ðɛt],[ðɛɾ]	[ðɨt], [ðɨ], [ðɨɾ]
I	[aɪ]	[ɑ],[ʌ],[æ]	[ə]
it	[ɪ],[ɪt],[ɪɾ]	[ʊt],[ʊ],[ʌ]	[ɨ],[ə],[ət]
you	[yu],[u],[yʉ]	[yɪ],[ɪ],[i]	[yɨ],[y],[ɨ]

Table 1. Most frequent pronunciations of the 10 words, grouped into basic, full, and reduced-vowel pronunciations. For each word we have shown the three most common tokens of each type of pronunciation in order of frequency.

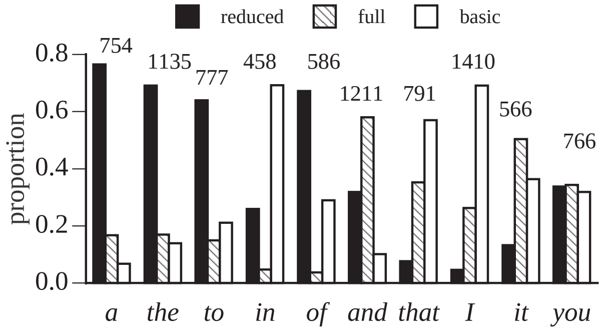

Figure 1. Proportion of basic, full, and reduced forms for the 10 function words. Total occurrences appear above.

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. RATE OF SPEECH

Speech researchers have long noted the association between faster speech, informal styles, and more reduced forms. (For a recent quantitative account of rate effects in Switchboard, see Fosler-Lussier and Morgan (1998)). 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. 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, a range which covers about 90 percent of the tokens, the estimated increase in the odds of full to reduced vowels was 2.2, i.e. the odds of a full vowel at the slow rate was 2.2 times the odds at the faster rate. Basic vowels also become more likely at slower rates, with an effect of about the same magnitude. These are both very highly significant effects ($p < .0001$).[3] Rate also did not affect all the words equally. The most strongly affected words were *a*, *the*, *to*, *and*, and *I*. Notably, regressions for *that* and *it* did not show rate effects for any of the three vowel or coda reduction measures.

4. PLANNING PROBLEMS

The production of speech is accompanied by a variety of disfluencies, whose characteristics have been extensively documented. In particular, it appears that some 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 are likely to cause words in immediately preceding speech to have less reduced pronunciations. They found this to be true for *the*, and suggested that the pronunciation [ði] is used by the speaker as a signal of impending problems in production. We follow earlier research in taking pauses, filled pauses like *uh* or *um*, and repetitions to be symptoms of planning problems. Each of the functors in our corpus was coded as belonging to a planning problem context if it was followed by one of these disfluencies.[4] Although the functors differed in their frequency of occurrence in the context of a planning disfluency, as can be seen in Table 2, this difference does not appear to significantly affect the reduction variables discussed below.

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 2. Percentage of occurrences of each word before a disfluency.

The effect of a planning problem on word length was massive and across-the-board (see Figure 2). The effect, both overall and for each word, remains after partialling 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. (Our regression models also included significant interactions between disfluency and rate and between disfluency and predictability.)

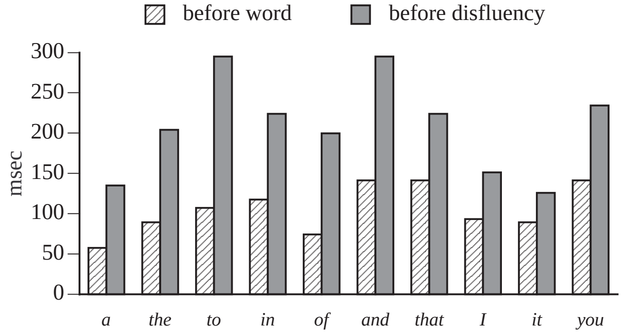

Figure 2. Average length for function words when followed by another word or by a disfluency.

A following disfluency also strongly affects the other reduction measures. Overall, the odds of a full versus a reduced vowel was increased 3.5 times before a disfluency. However, compared to the effects on length, the effects on reduction measures were uneven over the different words. As can be seen in Figure 3, the strongest effects on full/reduced are found for *to*, *I*, *a*, and *the*, and

---

[3] There was a significant interaction between rate and disfluency (stronger disfluency effects at slower rates), which we have ignored in reporting effects, since it did not affect their magnitude greatly.

[4] We were unable to code other symptoms of repair such as cutoffs, restarts, and editing phrases (e.g. *I mean*). Thus we would perhaps incorrectly interpret a cutoff followed by a pause as a symptom of a planning problem rather than as initiating repair of previous speech.

the weakest ones for *it* and *you*.[5] (Individual regressions on the latter two did not show a significant effect.) Only *a*, *the*, *to*, and *you* showed strong effects for the basic/full measure, with *and* and *I* showing smaller significant effects. The odds of a coda obstruent being present were about 5 times greater before disfluencies for *and* and *of*, but no effect was found for *it* and *that*. Overall, *a*, *the*, and *to* are the most sensitive to disfluencies, followed by *and* and *I*.

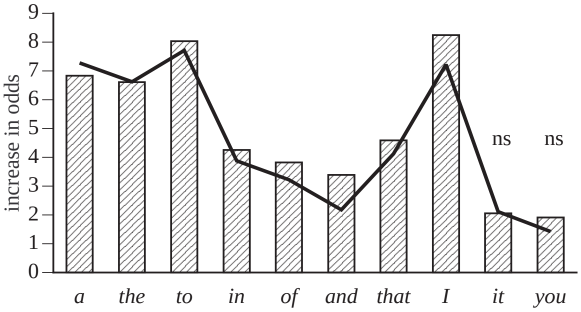

Figure 3. Effect of disfluency on the odds of a full versus a reduced vowel. The bars represent observed values. The line represents the increase in odds estimated from a regression model. Thus the line chart partials out the other variables (rate of speech and predictability).

5. FOLLOWING CONSONANT/VOWEL

A general fact about weakening processes is that the form of a word is influenced by the segmental context — in particular, more reduced forms tend to occur before a consonant than before a vowel (Rhodes 1996, inter alia). This may result in an allophonic effect such as the widely studied loss of final t and d (Neu 1980, inter alia). Alternatively, it may be an allomorphic one, as in the case of *the* with [ði] before vowels alternating with [ðə] before consonants (Keating et al. 1994).

Indeed, we found significantly less reduction in all four variables when the next word began with a vowel than when it began with a consonant. Table 3 shows the effects for individual words. For length, the values are the factor by which the duration of the word is longer before a vowel than before a consonant. For the other variables, the values are the increase in the odds of the first category before a vowel.[6] The effect was uneven across the ten words. 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 that an allomorphic account of a [tu]/[tə] or [ʌv]/[ə] alternation might be entertained.

	a	*the*	*to*	*in*	*of*	*and*	*that*	*I*	*it*	*you*
length	1.5	1.2	1.3	1.1	1.5	1.1	1.4	1.1	1.7	ns
full/red	9.1	14	4.3	ns	1.7	1.3	ns	ns	ns	$\boxed{\mathbf{.5}}$
basic/full	ns	8.8	6.6	ns	17*	ns	ns	ns	6.1*	ns
coda	—	—	—	—	17	ns	9.1	—	8.3	—

Table 3. Estimated effects of a following vowel (versus consonant) on reduction variables.

6. PREDICTABILITY AND FREQUENCY

The observation that higher-frequency words are more likely to have weakened pronunciations goes back at least to the late 19th century. Jespersen (1923), in commenting on this, emphasizes that frequency alone can be misleading, if the predictability of the word in its context is not also taken into account. We found no affect of word frequency on any of our measures of reduction. This was probably because there was relatively little difference in frequency among these most frequent words. 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.

	a	*the*	*to*	*in*	*of*	*and*	*that*	*I*	*it*	*you*
length	1.3	1.2	1.3	ns	1.3	1.2	ns	ns	1.4	ns
full/red	ns	ns	ns	3.2	ns	4.8	ns	$\boxed{\mathbf{.2}}$	ns	$\boxed{\mathbf{.2}}$
basic/full	ns	ns	ns	ns	8	ns	ns	ns	12	$\boxed{\mathbf{.1}}$
coda	—	—	—	—	2.3	ns	ns	—	ns	—

Table 4. Estimated effects of greater unpredictability on reduction.

In general, greater predictability increases the likelihood of reduction. Table 4 summarizes the effects, giving the factors of increase in length when it is highly unpredictable (ln p = -2.25) than when it is highly predictable (ln p = -.25). For the categorical variables, the values are the increase in the odds of the first category for highly unpredictable tokens. Length was the variable most affected by predictability, with an overall significance of $p < .0001$, but the effect was not significant for all words. In the expected direction, *in* and *and*, for full/reduced vowel, and *of*, for coda, showed strong effects ($p < .001$). We were surprised to find effects in the opposite direction for *you* and *I* (boxed values in Table 4). In predictable contexts they were both more likely to be full and *you* also was more likely to have a basic vowel.

7. COLLOCATIONS

In an attempt to understand the significant effect of predictability on the coda consonant of *of*, and the inverse effects of predictability on length and vowel quality of *you*, we looked at the lexical and syntactic context.

Much of the explanation for the greater reduction of *you* in unpredictable contexts came from the phrase *you know*. The 47% (359/766) of the instances of *you* occurring in *you know* were less syntactically predictable and more likely to be reduced than other instances of *you*: 47% vowel reduction for *you know* versus 22% for other instances of *you*. Excluding the instances of *you know* cut the effect of predictability on reduction in half. But there still remained an inverse effect of predictability on reduction. Another 12% of the instances of *you* occurred in an aux-inversion construction (*do you*, etc). These were extremely predictable from context but were statistically equally likely to be reduced as other *you*s. This might suggest that something about the aux-inversion context selects for non-reduced forms (perhaps the effect of focus in questions), counterbalancing the effect of predictability.

Collocational effects also explain much of the effect of predictability on reduction of *of*, which was significantly more likely ($p < .001$) to have no coda in predictable partitive constructions (*kind of*, *lots of*, etc) than in other uses (such as *thought of*, *outside*

---

[5] The values for *I* are unreliable because of a small cell count.

[6] Starred values are unreliable because of some low cell counts.

of). This suggests that the partitive construction may be stored or unitized as a mental routine.

8. ADDITIONAL EFFECTS

Following suggestions from research on repair (Fox and Jasperson 1995, inter alia) that planning problems may tend to be located early in turns, we examined the effect of word position on reduction and on the likelihood of disfluencies. No effect was found for either. The hybrid nature of our pseudo-utterances may be one reason for this; it may also be necessary to control for additional factors such as turn length, structure, and function.

The prosodic context provided by the following word can also be expected to influence reduction variables (e.g. the lengthening rule proposed by Bolinger 1986). Whether the next word's initial syllable contains a full or reduced vowel does affect length of the function word ($p = .0007$), whether its vowel is full or reduced ($p = .0006$), and whether its vowel is basic or full ($p = .005$), but not the presence of a final obstruent coda. The effects interact strongly with the presence or absence of a consonantal onset in the next word, appear to differ across words in complex ways, and are in general somewhat weaker than the effects discussed above. They require further study, especially with regard to possible relations with effects from stress and intonation.

9. CONCLUSIONS

Our results show that planning problems, predictability, segmental context, and rate of speech all play strong and independent roles in whether a word is reduced, for all measures of reduction. Planning problems are by far the strongest factor, despite the coarseness of our disfluency-based metric.

There are intriguing differences among the ten words in how they are affected by the various factors, some of them confirming earlier research, such as the preference for deletion of final voiced obstruents versus voiceless ones. It is possible that the strong effects of planning problems on the articles *a* and *the* and the pronoun *I* have something to do with their common occurrence at the beginning of turns, while accusative pronouns are known not to be subject to certain kinds of repair (Clark and Wasow in press). In addition, the apparently anomalous behavior of *you* and *I* in predictable contexts clearly deserves more attention, given their similarity in form and function. But in general, we have not focused on the individual differences in this paper.

Our results suggest that lexical representations of individual words may be more numerous than models of speech production have usually assumed (e.g. perhaps *to*, *of*, and *and* in addition to the more commonly noticed *the* and *a*), and furthermore that their selection is sensitive to a wide range of factors, notably the activities of monitoring and repair. Integrating the effects of rate, style, segmental context, and prosodic context on the durations and forms of the word, is also readily compatible with the models and concepts of gestural phonology (Browman and Goldstein 1992). Another important factor that we hope to address in future work is part of speech; for example reduction is more common in the complementizer *that* than the pronoun *that*, etc. We are also examining more sophisticated measures (non-retrospective predictability, change in pronunciation of word through a conversation) to investigate the difference between speaker-centered explanations for the predictability effect (Bybee 1996) and a hearer-centered explanations (Fowler and Housum 1987).

Our results also have important implications for automatic speech recognition. While C-onset versus V-onset effects can be captured to some extent by current triphone pronunciation models, the other factors studied here are not. Planning problems could be handled with simple modifications such as repetition-detection and the use of a silence phone. Speaking rate and word predictability would require the recognizer to change pronunciation models dynamically as these factors change. We feel that these are promising directions for future investigations of ASR pronunciation models.

Acknowledgements

This project was partially supported by NSF IIS-9733067, NSF SGER IRI-9713346, and the Center for Language and Speech Processing at The Johns Hopkins University. Many thanks to Elizabeth Shriberg, Steve Greenberg and two anonymous reviewers.

References

D. Bolinger. 1986. *Intonation and its parts: Melody in spoken English*. Stanford University Press, Stanford.

C. P. Browman and L. Goldstein. 1992. Articulatory phonology: An overview. *Phonetica*, 49:155–180.

J. L. Bybee. 1996. The phonology of the lexicon: evidence from lexical diffusion. In M. Barlow and S. Kemmer, editors, *Usage-based Models of Language*.

H. H. Clark and T. Wasow. in press. Repeating words in spontaneous speech. *Cognitive Psychology*.

E. Fosler-Lussier and N. Morgan. 1998. Effects of speaking rate and word frequency on conversational pronunciations. In *ESCA Tutorial and Research Workshop on Modeling pronunciation variation for automatic speech recognition*.

C. A. Fowler and J. Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26:489–504.

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*, pages 77–134. John Benjamins, Amsterdam.

J. E. Fox Tree and H. H. Clark. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62:151–167.

J. Godfrey, E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, pages 517–520.

S. Greenberg, D. Ellis, and J. Hollenback. 1996. Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *ICSLP-96*, Philadelphia.

O. Jespersen. 1923. *Language*. Henry Holt, New York.

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.

H. Neu. 1980. Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In W. Labov, editor, *Locating Language in Time and Space*, pages 37–54. Academic, New York.

R. A. Rhodes. 1996. English reduced vowels and the nature of natural processes. In B. Hurch and R. A. Rhodes, editors, *Natural Phonology: The State of the Art*, pages 239–259. Mouton de Gruyter.