Effects of disfluencies, predictability, and utterance position on word form variation in English conversation

Alan Bell and Daniel Jurafsky
Department of Linguistics, University of Colorado, Boulder, Colorado 80309-0925

Eric Fosler-Lussier
Bell Laboratories, Lucent Technologies

Cynthia Girand
Department of Linguistics, University of Colorado, Boulder, Colorado 80309-0925

Michelle Gregory
Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island

Daniel Gildea
University of Pennsylvania, Philadelphia, Pennsylvania

Function words, especially frequently occurring ones such as (*the, that, and*, and *of*), vary widely in pronunciation. Understanding this variation is essential both for cognitive modeling of lexical production and for computer speech recognition and synthesis. This study investigates which factors affect the forms of function words, especially whether they have a fuller pronunciation (e.g., ði, ðæt, ænd, ʌv) or a more reduced or lenited pronunciation (e.g., ðə, ðɨt, n, ə). It is based on over 8000 occurrences of the ten most frequent English function words in a 4-h sample from conversations from the Switchboard corpus. Ordinary linear and logistic regression models were used to examine variation in the length of the words, in the form of their vowel (basic, full, or reduced), and whether final obstruents were present or not. For all these measures, after controlling for segmental context, rate of speech, and other important factors, there are strong independent effects that made high-frequency monosyllabic function words more likely to be longer or have a fuller form (1) when neighboring disfluencies (such as filled pauses *uh* and *um*) indicate that the speaker was encountering problems in planning the utterance; (2) when the word is unexpected, i.e., less predictable in context; (3) when the word is either utterance initial or utterance final. Looking at the phenomenon in a different way, frequent function words are more likely to be shorter and to have less-full forms in fluent speech, in predictable positions or multiword collocations, and utterance internally. Also considered are other factors such as sex (women are more likely to use fuller forms, even after controlling for rate of speech, for example), and some of the differences among the ten function words in their response to the factors. *© 2003 Acoustical Society of America.* [DOI: 10.1121/1.1534836]

PACS numbers: 43.70.Bk, 43.70.Fq [AL]

I. INTRODUCTION

The modern availability of large online labeled corpora of conversational speech is a boon to the researcher studying phonological production. An obvious benefit of online conversational data is their ecological validity. But a less obvious benefit is the opportunity it affords for greatly expanding the range of situational and contextual effects that can be studied. Previous studies on read speech or reiterant speech, for example, have been able to study in detail the effect of phonetic variables such as segmental context on phonological variation. A number of variables, however, have received much less attention in earlier studies. In particular, the role of larger contexts such as prosodic context, lexical context, and the environment of the production task, is much less clear, particularly in natural conversational settings, and particularly for disfluent speech. It is essential to understand the role of these contextual factors in order to inform models of speech production.

This study investigates how the forms of English words in natural conversation are systematically affected by three such contextual variables: the presence or absence of neighboring disfluencies, the predictability of the word from the neighboring lexical context, and the position of the word in utterances. More specifically, it is hypothesized that words have stronger, less lenited forms in the presence of disfluencies, when they are less predictable, and when they occur at the beginning or end of utterances.

The first of these factors concerns a ubiquitous aspect of the production process itself, namely the disfluencies that arise when a hitch occurs in the flow from concept to speech. Previous studies have suggested that the surface form of words seem to be different when the speaker is experiencing lexical production planning problems. For example, Fox Tree and Clark (1997) showed that the word *the* was more likely to be pronounced with a full vowel rather than a schwa in disfluent contexts (when followed by a pause, filled pause, or

repetition). Our goal is to extend the Fox Tree and Clark (1997) study by examining whether such planning problems affect words other than *the*. We also study the nature of the form variation itself.

The second factor in our study is contextual predictability. Frequency and predictability have played a fundamental role in models of human language processing for well over a hundred years (Schuchardt, 1885; Jespersen, 1922; Zipf, 1929). But while modern models of human language comprehension often assume that probabilistic information plays a role in the access and disambiguation of linguistic structures (Jurafsky, 1996; MacDonald, 1993; McRae *et al.*, 1998; Trueswell and Tanenhaus, 1994), the role of probability in production is much less well understood. It is known that frequent words are shorter and more often reduced or lenited (Zipf, 1929; Fidelholz, 1975; Rhodes, 1992, 1996), that a second mention of a word is shorter than the first mention (Fowler and Housum, 1987), and that words which are more contextually predictable are produced in a less intelligible manner (Lieberman, 1963). In earlier work, (Jurfsky *et al.*, 2001; Gregory *et al.*, 1999) we proposed the *Probabilistic Reduction Hypothesis* to link these phenomena: word forms are reduced when they have a higher probability. The probability of a word is conditioned on many aspects of its context, including neighboring words, syntactic and lexical structure, semantic expectations, and discourse factors. In this paper we examine the role of local lexical probability: the probability of a word given the neighboring word or words. Our goal is to understand how this kind of local probabilistic context affects surface phonological and phonetic form, and how it relates to other kinds of context. We also ask whether the influence of a word's predictability is limited to the selection of alternate wordforms during lexical access, or whether predictability also influences surface phonetic form directly.

The third contextual factor we investigate is prosodic structure. The location of a word in larger prosodic domains such as utterances, turns, intonational phrases, and phonological phrases plays an important role in reduction. Studies of language change and of pronunciation variation have long accepted three main effects—final lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*), initial strengthening (i.e., more extreme articulation) (Fougeron and Keating, 1997; Byrd *et al.*, 2000, *inter alia*), and final weakening (i.e., less extreme articulation). During the last several decades more and more quantitative studies have helped make our understanding of these general effects more precise; see Fougeron and Keating (1997) for a review. Many of these results, however, derive from laboratory paradigms like reiterant speech, and have not been tested on natural speech production. Furthermore, it has been difficult to tease apart prepausal lengthening from lengthening at the edge of prosodic domains. We attempt to address these questions in the domain of natural conversational speech production.

How shall we investigate the effect of these factors? Natural speech corpora offer a number of potential dependent variables to use to study variation in phonological production. Previous research suggests that lenition and reduc-

tion, or alternatively, lengthening and strengthening, are associated with context. We therefore focused on this dimension of variation, selecting three dependent factors: duration of the entire word, categories of vowel quality, and presence or absence of coda obstruents. Longer pronunciations, with citation vowels or full vowels, are more frequent in explicit (e.g., formal, lento) styles; shorter pronunciations, with reduced or elided vowels and/or elided consonants, are more frequent in elliptical (e.g., casual, allegro) styles. These three variables thus reflect a scale of lenition, weakening, or reduction. For convenience we will use the term "reduced" throughout this paper to refer to the more elliptical forms. Other aspects of reduction, such as elision of initial consonants or consonant weakening, were not considered.

We investigate this reduction or lenition not in every word, but only in ten of the most frequent English words, namely the function words *I*, *and*, *the*, *that*, *a*, *you*, *to*, *of*, *it*, and *in*. Why is the study limited to just these words? Briefly, there were three main reasons. A study covering all words was judged too ambitious and too complex for an initial application of the multidimensional analysis methods to be used, and one must start somewhere. The high frequencies of occurrence of these words, their especially great form variation, and their common monosyllabic form offered important advantages to the analysis. The fact they are also function words, that is strongly associated with syntactic and semantic/pragmatic structures, was not a primary consideration. Finally, and crucially, the fact that such words are not usually accented allowed us to avoid problems of controlling for the interaction of segmental form and presence of accent. If the contextual effects on reduction that we postulate exist, there should be strong evidence for them in the most frequent words; the possibly more difficult task of verifying that the effects also hold throughout the lexicon can be left for further research.

Our data is drawn from the Switchboard corpus of telephone conversations between strangers collected in the early 1990's (Godfrey *et al.*, 1992). We chose the Switchboard corpus for our research because various portions of it have been phonetically transcribed, coded for part of speech, syntactically parsed, and segmented into utterance-like units.

The next section of the paper, Sec. II, summarizes our methodology for extracting and coding forms, and analyzing form variation. Section III then describes details of the various control variables; rate of speech, phonetic context, pitch accent, etc., and summarizes their effects. Section IV focuses on our first contextual variable, the presence of disfluencies, which we take to be largely associated with problems in planning speech. Section V focuses on the second contextual variable, word predictability from neighboring words. Section VI deals with the last contextual variable, the position of a word in prosodic domains. Section VII concludes with a discussion of the results and their implications.

II. METHODOLOGY

A. The corpus

As described above, our observations of the ten function words *I*, *and*, *the*, *that*, *a*, *you*, *to*, *of*, *it*, and *in* were drawn

from the phonetically transcribed portion of the Switchboard corpus collected in the early 1990's (Godfrey *et al.*, 1992). The corpus contains 2430 conversations averaging 6 min each, totaling 240 h of speech and approximately 3 million words. The corpus was collected at Texas Instruments, mostly by soliciting paid volunteers who were connected to other volunteers via a robot telephone operator, and was then transcribed by court reporters into a word-by-word text.

Approximately 4 h of this speech was phonetically hand-transcribed at ICSI (the International Computer Science Institute) by linguistics students at UC Berkeley (Greenberg *et al.*, 1996; Greenberg, 1997) as follows. The speech files were automatically segmented into "fragments" at turn boundaries or at silences of 500 ms or more. The transcribers were given these strings, the word transcription, and a rough automatic phonetic transcription which was automatically aligned to the wavefile at syllable boundaries. They then corrected this rough phonetic transcription, using an augmented version of the ARPAbet. The transcribers also corrected the syllable boundary marks and the silence onsets and offsets. In general, transcribers were instructed to pay careful attention to both the waveform and spectral displays of the signal in making their decisions. In cases where no specific event could be found to mark a syllable boundary, guesses were made using tables of the duration distributions of particular segments. These boundary marks were then used to automatically compute syllable durations. Similarly, pause durations were computed for portions of the signal not attributed to a syllable. The hand-labeled and hand-segmented syllables were then automatically aligned against the word transcription, resulting in a duration for each word. Since the current study only considers monosyllabic words, in many cases these durations correspond exactly to the hand-labeled syllable boundaries. In some cases where resyllabification occurred, the automatic alignment did slightly shift the boundaries. The entire corpus contains roughly 38 000 transcribed word tokens.

Approximately two-thirds of this phonetically transcribed corpus (henceforth the ICSI corpus) was also part of the utterance-segmented portion of the Treebank III release of the Switchboard corpus (Marcus *et al.*, 1999). In this release, 1155 of the 2430 conversations were segmented by the Linguistic Data Consortium (LDC) into approximately the 205 000 utterance-like units described in Sec. VI (Meteer *et al.*, 1995).

Our database thus combines information from three sources: the original lexically transcribed Switchboard corpus, the Treebank III utterance segmentation, and the ICSI phonetically transcribed corpus. All three of these corpora, together with documentation describing them, are available from the Linguistic Data Consortium at `http://www.ldc.upenn.edu/`

From the phonetically transcribed data, we extracted 9926 occurrences of the ten function words. We immediately eliminated 801 occurrences whose surface form clearly indicated an alignment error or a transcription error, such as the word *you* pronounced [rju], or the word *you* pronounced [ði]. This left 9125 tokens of the ten function words. Of these, 404 were alternate forms such as *an, I'd, I'm, I'll,* and *you'd,*

TABLE I. Most frequent pronunciations of the ten words, grouped into basic, full, and reduced-vowel pronunciations. For each word the three most common tokens of each type of pronunciation are listed in order of frequency.

	Basic	Other full	Reduced
a	[eɪ]	[ʌ],[ɪ]	[ə],[i]
the	[ði],[i],[di]	[ðʌ],[ðɪ],[ʌ]	[ðə],[ði],[ə]
in	[ɪn],[ɪ],[ɪr],	[ɛn],[ʌn],[æn]	[in],[n̩],[ən]
of	[ʌv],[ʌ],[ʌvv]	[ɪ],[i],[ɑ]	[ə],[əv],[əf]
to	[tu],[tʉ],[ru]	[tʊ],[tɪ],[tʌ]	[tə],[ti],[ə]
and	[æn],[ænd],[ær]	[ɛn],[ɪn],[ʌn]	[in],[n̩],[ən]
that	[ðæ],[ðæt],[æ]	[ðɛ],[ðɛt],[ðɛr]	[ðit],[ði],[ðir]
I	[aɪ]	[ɑ],[ʌ],[æ]	[ə]
it	[ɪ],[ɪt],[ɪr]	[ʊt],[ʊ],[ʌ]	[i],[ə],[ət]
you	[yu],[u],[yʉ]	[yɪ],[ɪ],[i]	[yi],[y],[i]

you're, etc., which because of their small numbers and incomparable forms, were excluded from our analysis. We also excluded 361 items which were coded as "nulls," i.e., as having no segmental realization except possibly as a featural modification of an adjoining word. (The discussion below on coding of vowel quality comments further on the null items.) This left 8362 items as input to our analyses. The actual sample sizes of most analyses are smaller than this, because not all variables apply to all the data or could not be defined for all the data; see the discussions below.

B. How forms were coded

The three dependent factors of duration, vowel quality, and coda presence were coded in the following ways.

(1) *Vowel quality*: We coded each vowel as **basic**, **other full**, or **reduced**. The basic vowel is the citation or clarification pronunciation, e.g., [ði] for *the*.[1] The reduced vowels are [ə] (arpabet [ax]), [ɨ] (arpabet [ix]), [ɚ] (arpabet [axr]), and [ɵ] (midcentral reduced vowel with more [o]-like or [u]-like coloring than [ə], not in the arpabet). Any other vowel is a full vowel. This three-way distinction is split into two binary contrast variables: full/reduced (basic and other full vowel versus reduced vowel) and basic/full. See Table I for the most frequent tokens of the words in each of the vowel quality categories.

(2) *Coda obstruent*: For words which have coda obstruents (*it, that, and, of*), we coded whether the consonant is present or not. The sonorant nasal codas of *in* and *and* were not considered.

(3) *Length*: We coded the duration of the word in milliseconds.

In general we relied on the ICSI transcriptions for our coding, using software to automatically assign a category to a transcribed word. Thus, for example, if the ICSI transcription of a word was [əv], our software automatically categorized the observation as *reduced vowel* and *coda present*. We judged the interlabeler agreements of the ICSI transcribers, reported between 72.4% and 76.9%, to be quite acceptable for this task. We did, however, check the data several ways, deleting or modifying some items. As mentioned above, we first examined every pronunciation of every word, and elimi-

nated 801 incorrect pronunciations that were due to alignment errors in our automatic word-segmentation program. We then listened 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 from these five classes were recoded, mainly [ux] as either a nonreduced high front round vowel [ʉ], as prescribed, or reduced [ɵ]; and [er] as either full [ɝ] or reduced [ɚ]. Some items were removed, mainly those transcribed as having no segment ("nulls"), since from our sample we judged that many were equally segmental as other transcriptions. Most of the incorrect coding of these words as having "no segments" was due to a mismatch between incorrect word transcriptions and the phonetic transcription for the utterances. In these cases, the phonetic labelers transcribed the utterance correctly but did not correct the original word-level transcription. The mismatch between these two produced a number of alignment errors which we eliminated.

Our judgments 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 43 reduced vowels. This suggests that there may be a bias toward full vowels in the transcription.

Neither we nor the original Switchboard Transcription Project at ICSI computed interlabeler agreement statistics for syllable duration labeling. We did, however, check some segmental durations, and while in many cases we might have slightly moved segment boundaries, we found no reason to believe there were any gross systematic errors in duration labeling.

The coding for each of the three major independent variables (planning problems, predictability, and utterance position) is described in the later sections pertaining to each variable.

C. Controlling for possible confounds: Regression analysis

While the use of natural conversational corpora provides the benefits of situational validity and allows us a larger contextual window, it also presents a problem. Natural speech has myriad confounding factors that affect form variation such as phonetic factors, rate of speech, pitch accent, and sociological factors like age and sex. These factors are typically correlated. We use multiple regression, both linear and logistic, to examine the individual contributions of a variable in this situation.

A regression analysis is a statistical model that predicts a *response variable* (in this case, the word duration, or the frequency of vowel reduction) based on contributions from a number of other *explanatory factors* (Agresti, 1996). 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 explanatory variables, adding the explanatory variable in question

produced a significantly better account of the variation in the response variable.

For duration, which is a continuous variable, we use ordinary linear regression. For vowel quality, and coda presence, which are categorical variables, we use logistic regression. Logistic regression models the effect of explanatory variables on a categorical variable in terms of the *odds* of the category, which is the ratio $[P(\text{category})]/[1-P(\text{category})]$. 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 17 percent of the time, and with a reduced vowel, 83 percent; the odds of a full vowel are $17/83 = 0.20$ (to 1).

It is important to understand that the goal of the regression analyses is not to create a model that will predict the forms of function words. It is primarily used as a tool to evaluate the significance and magnitude of selected factors in the presence of other correlated factors, possibly also significant.

Of course, establishing that a factor can contribute additional improvement to a model is one of the basic facts needed to construct production models. Much more, such as details of dependencies among factors and magnitudes of effects at high and low values of factors, is also needed. Some selected questions of this sort that appear to be particularly important are explored in the sections below. For example, we generally report important interactions, notably the greater effect of predictability from a preceding word for more frequent word combinations (Secs. V A 1, V A 2). Hypotheses about certain factor dependencies are tested with specific regression models, e.g., relations between disfluencies and utterance-initial position, Sec. IV B 1; and relations between word duration and vowel reduction (Secs. III A, IV A, V A 3). A few comparisons between alternative models are tested, e.g., the comparison between a two-factor model distinguishing preceding and following disfluencies and a single-factor model which does not, Sec. IV A.

The size of a factor's effect is of considerable importance, since a factor can be a significant addition, but have a relatively small effect. The level of significance of an effect is often associated with its magnitude—an effect significant at $p = 0.0001$ is likely to be greater than one that is significant at $p = 0.01$. This is not a generally appropriate measure of effect magnitude, however, so two other measures are commonly used. One is based on the estimated weight of the factor in the regression equation; the other is based on the proportion of the total variation that the factor accounts for. The weight-based measure, which is the more direct of the two, is reported for the main results. It is a ratio derived from two parameters—the estimated weight and the range of the factor. In the simplest case for a categorical factor like presence of a disfluency, the range is 1, so that the effect magnitude is simply proportional to the regression weight. In Sec. IV A, the effect of a disfluency on vowel reduction is reported as 1.68, meaning that all other factors being equal, in a disfluent context, the estimated odds that the word contains a full vowel are 1.68 times the odds of a full vowel in a fluent context. This value is calculated by taking the regression coefficient of the disfluency factor as a power of 10,

since the regression equation is based on log odds. For continuous factors, a range representing the middle 90 percent of the data is used, from the 5th to the 95th percentiles. Thus, in Table X, the magnitude of the effect on duration of the conditional probability given the previous word, 0.80, means that the estimated duration of the most predictable words (at the 95th percentile) are 0.80 times shorter than the least predictable words (at the 5th percentile).

One of the assumptions of regression analyses is that the items in the data are independent. This assumption is surely violated to some extent by our data, since many of the same items are uttered by the same speaker, or in the same conversation. A more serious violation occurs when two words are adjacent. Just how to best deal with this inherent weakness of corpus studies is not clear. Sampling one item from each conversation, or part of a conversation, was judged too costly. It would drastically reduce the power of the analyses and their generality. One reason for examining the ten most frequent function words was the expectation that in most instances such words would be separated, and occur in separate phrases. Although this is usually the case, about 20 percent of the items do occur adjacently in combinations such as *of the* and *that I*, which is not very surprising just given their high frequency of occurrence. The consequence of the non-independence of such items is that the significance values are inflated to some extent. It is thus recommended that the reader not take the reported levels literally, but as an informed indicator of the relative significance of an effect. Where the significances are very great, this is of little concern, but becomes more of one for more marginal ones. While we have reported some effects at levels up to the conventional 0.05 level, it seems prudent to regard any result above $p = 0.01$ as marginal.

The results are of course subject to the usual limitations of such analyses, most notably that they apply strictly only to the present database and to the particular operational coding used. In many ways, the database can be considered generally representative of American English conversation. But some of the specific characteristics of the data, for example, the particular way that fragments of conversations were selected for the ICSI database, require simplifications in variable definitions and sample selections that inevitably introduce some degree of bias. Examination of many such cases has not yielded any reason to think that the distortions are large enough to invalidate the main results. Nevertheless, it is perhaps well to regard the quantitative measures of the results as pertaining to this database, and to take the results more qualitatively as a basis, together with further research, for constructing production models.

III. CONTROL FACTORS

The reduction variables are each influenced by multiple factors that must be controlled to assess the contribution of the explanatory variables—presence of disfluencies, predictability, and position in turn. While it is of course not possible to control for every factor which influences reduction, we consider here the ones that are, from prior research, most likely to play a large role.
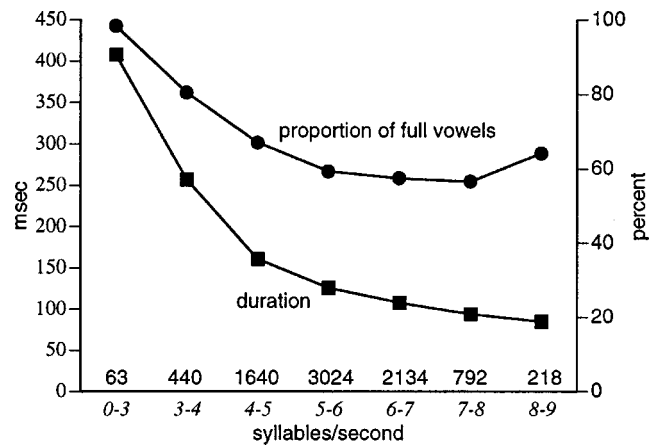

FIG. 1. Function word durations and proportions of full vowels by rate of speech. The scale for duration is on the left axis, the scale for full vowels is on the right. The number of observations for each rate category appears at the bottom of the graph.

(i)   rate of speech of the speaker in syllables/second;
(ii)   segmental context;
(iii) prosodic factors;
(iv)   age and sex of speaker and hearer; and
(v)   individual characteristics of the ten function words.

The focus of the paper leads us to regard these as control factors rather than object of study in their own right. The role of rate, phonetic context, and prosody in reduction is of course well-established. A detailed study of speaker and hearer effects in conversational speech is beyond the scope of this paper, in spite of its considerable interest. This section, therefore, reports primarily the details of the variables we selected to control these factors. Selected results about how these variables affect reduction are also presented.

A. Control factors: 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 (1999)]. We measured rate of speech at a given function word by taking the number of syllables per second in the speech fragment immediately surrounding the word, up to the nearest pause or turn boundary on each side. Fifty-one words with extremely slow or extremely fast rates were excluded from regression analyses. Unsurprisingly, rate of speech affected all measures of reduction. Words were more reduced when they were spoken more quickly. 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 is 2.2. That is, the odds of a full vowel at the slow rate is 2.2 times the odds at the faster rate. Figure 1 compares observed proportions (or averages, for length) with predicted values for five categories of rate along the range from 2.5 to 7.5 syllables per second. (The increased proportion of full vowels at the highest rate category is presumably not systematic.)

For all measures, there seems to be a limit effect for

faster rates; this is accounted for in the regression model by using log (rate) as the main explanatory variable as well as a (highly significant) quadratic $\log^2$ (rate) term. The overall effect of rate is weaker for coda deletion than for the other measures of reduction; in addition, the effect on deletion is largely confined to the slower rates.

Are the shortening effects for function words solely a consequence of a greater proportion of reduced (and shorter) vowels at faster rates? If they were, the apparently gradient effect of rate of speech on durational shortening might represent not a gradient effect of rate, but a categorical effect, stemming from more frequent selection of reduced vowel forms at faster speech rates. It turns out that there is a substantial additional shortening effect of rate even after accounting for vowel reduction and coda deletion. Overall, with no other variables involved, rate accounts for 17.9 percent of the variation in duration of the function words. With no other variables involved, vowel reduction and coda deletion account for 18.4 percent of the variation. After controlling for vowel reduction and coda deletion, rate still accounts for an additional 13.9 percent of the variation. A final characteristic of rate is that it did not affect all the words equally. The most strongly affected words were *a*, *the*, *to*, *and*, and *I*. Notably, regressions for *that*, *it*, and *in* did not show rate effects for any of the three vowel or coda reduction measures.

B. Control factors: Segmental context

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, and references therein). 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). The preceding segmental context is presumed to have much less influence.

Thus, for each of the function word tokens, we recorded whether the following word began with a consonant or a vowel. To account for an interaction between this following segment and the final obstruent consonant of the function word itself, we distinguished four separate contexts: V#V, V#CV, VC#V, and VC#CV. The nasals of *and* and *in* were treated as if they belonged to the nucleus, both because they can be expected to behave differently from the obstruents, and also because the interplay between vowel nasalization and nasal consonant shortening is not captured by the ICSI phonetic transcription.

In addition, the metrical strength of the following word or words can also be expected to influence reduction. Here, we attempted to capture some portion of this influence by coding each function word with a variable distinguishing whether the vowel of the following syllable is full or reduced. In general, since reduced vowels cannot be stressed or bear intonational accents, this variable may be regarded as mainly differentiating cases where the next potential prosodically strong syllable either follows directly or else one or more syllables later. A more direct effect is predicted by

TABLE II. Observed average durations and reduced vowel percentages of closed-syllable (VC) and open-syllable (V) function words before words beginning with consonants and with vowels.

	Word	Next word	Duration (ms)	Percentage of reduced vowels
consonant	VC	CV	132	33.7
follows	V	CV	102	45.6
vowel	VC	V	158	29.7
follows	V	V	128	33.0

Bolinger's (1986) lengthening rule, which states that a full vowel is lengthened if the next vowel is also full.

Observed average duration and the percentage of reduced vowels are shown in Table II for the four contexts. Before a consonant in the next word, words are shorter and are more likely to be reduced. These differences were assessed by regressions after controlling for rate effects. For both vowel reduction and shortening, the onset of the following word has a very strong effect. Overall, the odds of vowel reduction are 1.63 times greater before consonants than before vowels, and item durations are 0.79 times shorter before consonants than before vowels. The consonant–vowel effect on vowel reduction is stronger for open-syllable words than for closed-syllable ones; the effect for closed syllables is still highly significant ($p = 0.0005$).

The full-reduced status of the vowel in the next word affects open-syllable items, whose vowels are more likely to be reduced if the next word has a full vowel in its first syllable (whether it begins with a consonant or not). This is a moderately significant effect ($p = 0.007$, odds ratio of 1.43); there is no significant effect of the following vowel for closed-syllable items. Duration is also affected by the category of the vowel in the next word, but in a complex way. In the VC#CV and V#V contexts, there is little effect. Open-syllable items before consonants (the V#CV context) are shorter (by a factor of 0.82) if a full vowel follows, but closed-syllable items before vowels (the VC#V context) are shorter (by a factor of .84) if a reduced vowel follows.[2]

As with rate, shortening effects are still strong after controlling for vowel reduction. Overall, for example, the onset of the following word accounts for 4.1 percent of the variance; within reduced or full vowels, it still accounts for 3.6 percent of the variance. Individual analyses by item largely confirm the overall results for reduction and shortening. Only *you* for lengthening and *that* and *in* for reduction fail to show significant effects, which of course may be partially laid at the door of the smaller sample sizes.

C. Control factors: Intonational accent

One of the most important factors influencing an English word's pronunciation is whether it receives accent or not. Presence of accent is surely highly correlated with longer duration, lack of vowel reduction, and lack of elision, and likely has systematic associations with the presence of disfluencies, a word's predictability, and its position in the intonational phrase, the explanatory variables that are considered here. The most general way of accounting for its role in wordform variation is to regard it as one of the attributes of

a word's form together with its segmental attributes of duration, vowel reduction, etc., that is, as a response or observational variable. Desirable as this might be, it entails analytic complexities and model-theoretic assumptions that seemed premature at our present stage of knowledge. The alternative is to focus on the word's segmental form, and treat the prosodic status as an explanatory variable, part of the general context in which the word occurs, and one of the factors influencing the form of the word. Since intonational accent is not transcribed in the ICSI database, we could not examine its effects or control for them directly. One of the main reasons for studying high-frequency function words was that they are unlikely to be accented. It was our hope that the possible confound of accent with variables such as disfluency and predictability would be so infrequent that it would have little influence on their analysis. Fortunately, we have been able to verify this perhaps incautious hope, making use of two small accent-coded subcorpora from Switchboard. The first was a small portion of Switchboard that has been coded for accent under the direction of Shattuck-Hufnagel and Ostendorf, an alpha-release version of which they generously made available to us. The Shattuck-Hufnagel/Ostendorf corpus used a labeling scheme called POSH (Shattuck-Hufnagel and Ostendorf, 1999), a simplification of the TOBI prosodic labeling standard (Silverman *et al.*, 1992). In addition to the Shattuck-Hufnagel/Ostendorf corpus, we coded a very small subsample of Switchboard consisting of 120 words selected from the longest tokens of each function word; it was composed of 10 tokens of each function word, except for those which may be pronouns, *I* (20 tokens), *you* (15 tokens), and *that* (15 tokens).

The overlap between the Shattuck-Hufnagel/Ostendorf corpus and the most inclusive sample used in our analyses (8311 words) was 560 words. Of this set, 53, or 9.5 percent, were accented. (A larger proportion, 23 percent, were accented in our 120-word sample, presumably because of its heavy bias toward items most likely to be accented.) A majority of the accented words was either *that* (16) or *I* (15); with *and* (6), *you* (5), and *in* (4), they accounted for all but seven of the accented words. This concentration of accent on particular function words more or less agreed with our sample, in which only four functors had more than one accented token: *I*, 12 of 20; *you*, 7 of 15; *that*, 4 of 15; and *and*, 2 of 10. It appears that function words are indeed not likely to be accented, but some function words are much less likely than others to bear accent.

In order to determine whether the accent-coded data were representative of our entire database of phonetically transcribed words, we compared relative frequencies of the function words, rates of reduction, duration, rates of preceding and following disfluencies, and preceding and following conditional and joint probabilities, using chi-square or Fisher tests for the categorical variables and t-tests for the continuous ones. Since only one of the nine comparisons was even close to significant, the subset of accented-coded data appeared to represent the overall sample reasonably well. We also examined the association of accent with disfluencies and with predictability. This confirmed our expectation that accented words would be more likely to occur in disfluent contexts and that their conditional probabilities would on average be lower than unaccented words. Only for a previous disfluency, however, was the difference significant (one-tailed Fisher test, $p = 0.001$), perhaps because of the small sample.

The main question, of course, is whether the effects of the explanatory variables remain after controlling for pitch accent. We addressed this by examining only the 385 words without accent. (The accented words were too few to make including them in an analysis useful.) The details of the comparison of this analysis with the full analyses are presented in the following sections that treat the effects of disfluencies and of predictability on duration. Overall, as will be seen, the effects that are found for the unaccented word sample are similar to those for the overall sample uncontrolled for accent. These results are necessarily preliminary and incomplete. We did not examine whether accent might be masking the role of disfluencies and predictability on vowel reduction, basic versus nonbasic vowels, and coda deletion. There were not enough data to examine effects of position or effects for individual words. The clear results for duration, however, support our strategy of examining the factors affecting form variation in function words in the absence of controls for accent. Note also that the results for the individual words which virtually never receive accent are further support. Obviously, important questions about the role of accent remain, both for function words and content words.

D. Control factors: Age and sex of speaker and hearer

Studies of socially sensitive pronunciation variation such as the alternation of -*ing* and -*in* (Wald and Shopen, 1981) have shown that the status of speaker and hearer is often a factor in such variation. It is likely that such influences extend to our reduction variables, given that all our indices of variation are doubtless linked to the choice of elliptical versus explicit styles of speech, which is in turn sensitive to the speech situation. While an earlier study of the TIMIT corpus of read speech by Byrd (1994) did not find an effect of speaker sex on the duration of centralized vowels, she did find that men use certain more reduced forms such as taps and syllabic n more frequently than women. Previous research has also shown that rate and disfluencies are sensitive to the age and sex of speakers. Byrd (1994) found that men spoke TIMIT sentences on average 6.2 percent faster than women. Shriberg (1999), in her study of disfluencies in Switchboard, found that men had slightly more disfluencies per word than women. There is thus good reason *a priori* to control for speaker and hearer status. In this section we present a simple survey of the overall differences in reduction, rate, and disfluencies associated with the age and sex of speakers and hearers in our dataset.[3] Since this survey is meant only to provide a basis for the use of these factors as controls, we do not provide detailed analyses with individual assessments of significance. Some summaries of analyses for individual items are included in Secs. IV B, V B, and VI B. The more complex analysis needed to assess their effects on production is left for future study.
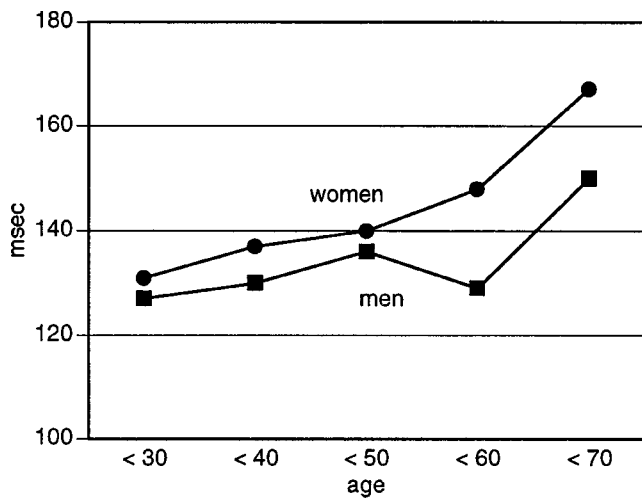
FIG. 2. Average word durations of function words of men and women speakers by age.

FIG. 3. Predicted average speech rates of men and women speakers by age.

The ages of the 497 participants in our sample of Switchboard conversations ranged from 18 to 68; the mean age of the speakers was 37. There were 191 men speakers and 172 women, and 237 men listeners and 216 women. More items were spoken by men (58 percent) than by women (42 percent).

1. Effects of speaker and hearer on reduction variables

All the reduction measures are affected by the speaker's status.

(i)   Duration: The average durations of function words are shown in Fig. 2 for men and women speakers by age category. Words spoken by women are longer (140 ms) than those spoken by men (131 ms), and words spoken by older speakers are longer (139 ms for speakers 40 and older versus 131 ms for those under 40), with the difference greater between older men and women.

(ii)   Vowel reduction: The sex of speaker has the strongest effect on vowel reduction; there is little difference for older or younger speakers. Words spoken by men are reduced 41 percent of the time on average, but only 34 percent of the time for women.

(iii)   Coda deletion: On the other hand, women speakers delete codas more frequently than men, 68 percent to 63 percent.

(iv)   Basic vowel: Basic vowels are used more by older speakers than by younger ones; speakers under 40 use basic vowels 60 percent of the time, but this increases to 66 percent for speakers 40 and older (69 percent for speakers 60 and older).

These uncontrolled differences are significant at levels from $p < 0.005$ to $p < 0.0001$.

Differences associated with listener status are much smaller, and not significant, except perhaps for vowel reduction. Words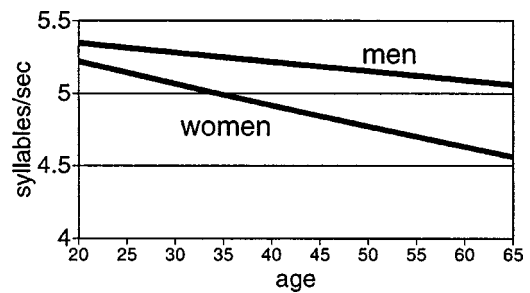 are more often reduced when spoken to younger listeners under 40 than to older listeners ($p < 0.05$). There were no dyad effects of speaker and listener age or of speaker and listener sex.

Overall, reduction in function words is affected mainly by the age and sex of the speaker, and mainly in the directions that one would expect from the usual correlations of speaker status and levels of formality in speech: longer durations with less reduction of vowels and greater use of basic vowels by women and by older speakers.

2. Effect of age and sex on rate and on disfluencies

On average men spoke 6.4 percent faster than women. Men had an average rate of 5.4 syllables per second: women, an average rate of 5.0 syllables per second.[4] Younger speakers spoke more quickly, 5.5 syllables per second for speakers under 30, compared to 5.1 syllables per second for speakers 50 and older. 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. These relationships are shown in Fig. 3, which presents the regressions of rate on age for men and women. There do not appear to be any differences in rate for different listener statuses.

The average rate of disfluency was 31.9 percent, where by disfluency we mean the presence of a disfluency either before or after a given function word. Interestingly, uncontrolled averages reveal little difference between men and women speakers of different ages, nor between men and women listeners of different ages. Since this was not consistent with Shriberg's (1999) results, we explored disfluency effects by controlling for rate and for the probability variables. The results agreed with Shriberg's finding that men have a higher rate of disfluency than women.

E. Control factors: Individual characteristics of the words

Different function words play different grammatical roles, have different distributions, have different kinds of meanings, and have different phonological forms. One should therefore expect some differences in how their reduction is affected by other factors. While it would be impractical and probably undesirable to control for item effects in analyses for overall effects of disfluencies, predictability, and utterance position, the idiosyncrasies of the ten function words unquestionably affect such results. It is thus important to compare their basic characteristics.

Bell *et al.*: Disfluencies, predictability, position, and form variation

TABLE III. Frequencies of occurrence of the ten function words in the data.

I	and	the	that	to	you	a	of	it	in	Total
1381	1203	1123	786	769	758	745	583	562	452	8362

First, some are more frequent than others in our sample, reflecting their relative frequency in conversational speech. *I*, *and*, and *the* are the most frequent, and *of*, *it*, and *in* are the least, as can be seen in Table III.

The frequency range from most to least frequent is about 3 to 1, quite modest for lexical frequency in general, but to be expected since these are the ten most frequent words in the Switchboard corpus. One consequence of this distribution is that one cannot investigate the effect of lexical frequency on reduction with this database. This is partly because of the narrow range of frequencies, but more crucially because item frequency is confounded with other item idiosyncrasies in this small set, and there is no way to pull them apart. The other issue is the relative influence of the items on the overall results. Clearly the most frequent words will have more influence than the least frequent ones, and this needs to be kept in mind in the following discussions. It is also not advised to view this as an improper distortion of the results, since after all, the proportions of each word reflect their relative occurrence in conversational speech.

Next, the items differ considerably in their average durations and average rates of occurrence of basic, full, and reduced vowels and of coda deletion, resulting in different base levels for the overall effects on these variables.

Figure 4 shows the average durations of the ten function words. In this and following figures and tables, the words have been grouped by dominant function: articles *a, the*; prepositions/particles *in, of, to*; conjunctions *and, that*; and pronouns *I, it, you*. *And* and *that* are notably longer, in part because their vowel is intrinsically long and because they have a complex syllable structure. Similarly, the shortness of the article *a* probably reflects its single vowel.

Striking differences in the average rates of occurrence of full, unreduced vowels can be seen in Fig. 5. Six words, including *and* 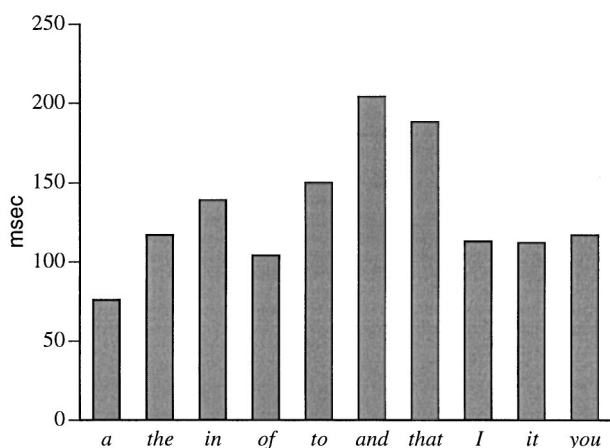and *that*, have relatively high proportions of unreduced vowels, between about 65 percent and 95 percent; the others, including the article *a*, are much less likely to have an unreduced vowel (between about 25 percent and 35 percent). Thus another reason that *and* and *that* are longer is because they more often have full vowels. Likewise, the frequent reduced forms of *a* contribute to its relative shortness. Obstruent codas are present in about the same proportion for *that*, *it*, and *of* (44 percent for *that* and *it*, 54 percent for *of*); on the other hand, the very infrequent presence of the final stop of *and* (14 percent) suggests that the alternation between [n] and [nd] may stem in part from selection between distinct lexical forms of *and*.

Such item differences can affect our results in two main ways. First, the longest and shortest (or most/least reduced) items may contribute to floor or ceiling effects for some factors. As we mention below in Sec. IV B, a ceiling appears to be at least one factor responsible for *that*, *I*, and *it* not showing fewer reduced vowels in the context of following disfluencies. Their proportion of unreduced vowels is already so high it cannot become much higher. A second way is for one or more of the items with atypical forms or behaviors to be disproportionately represented over the range of a factor. One example of this is the very frequent occurrence of *and* in utterance-initial position, discussed below in Sec. VI B. Since *and* is long, it should exaggerate an initial lengthening effect; as we see later, if *and* is excluded from the analysis, the effect on duration in initial position is indeed reduced, although it remains significant.

There are of course additional differences in the behavior of the function words with respect to disfluencies, predictability variables, and utterance position, which largely reflect their functional differences. Some of the most salient such differences are discussed in the following sections. In lieu of controlling for item differences, we note below the consistency of effects over the function words, or lack of it. This provides a general indication of the robustness of the effects, and in some instances of markedly aberrant items, it suggests certain factors which may be responsible.

FIG. 4. Observed average durations of function words. The average duration of all words is 135 ms.
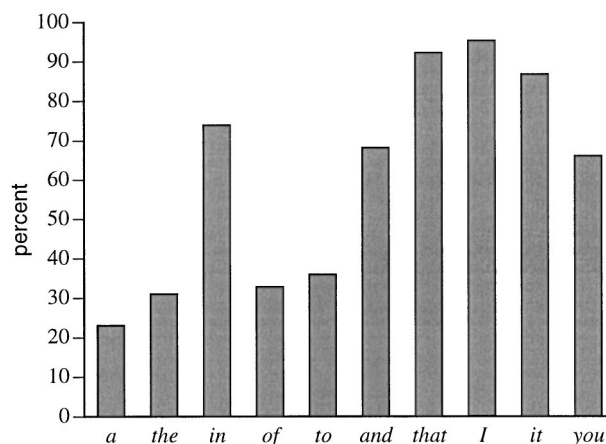
FIG. 5. Observed average frequency of occurrence of unreduced (full) vowels in function words. The average frequency for all words is 0.62.

TABLE IV. Observed durations, frequencies of basic and full vowels, and frequencies of coda presence for function words in fluent and disfluent contexts. The number of observations of the context categories appears in parentheses. Basic vowel frequencies are based on the 4886 words with full vowels. Obstruent coda presence frequencies are based on the 2947 words *and, it, of*, and *that*.

Context	Duration	Full vowel	Basic vowel	Coda presence
Fluent	109 ms (5480)	54% (5480)	64% (2936)	33% (1948)
Any disfluency	187 ms (2519)	77% (2519)	64% (1950)	39% (999)
Disfluency before	137 ms (1295)	73% (1295)	59% (940)	26% (366)
Disfluency after	222 ms (927)	80% (927)	66% (741)	42% (483)
Disfluency both	295 ms (297)	91% (297)	75% (269)	59% (150)

F. Summary of control factors

Most of the factors discussed above are controlled in our regression analyses by including appropriate variables in a base model. Our base regression models thus include the following variables.

(i)   Log rate of speech and log squared of rate of speech;
(ii) Syllable type of target (open, closed);
(iii) Whether initial segment of next word begins with consonant or vowel;
(iv) Whether following vowel is reduced; and
(v)   Age and sex of speaker, age of listener.

Included also were significant interaction variables, e.g., rate×speaker age. Some of these variables were dropped when they had negligible effect, e.g., listener age for duration analyses. The results presented below are based on analyses that controlled utterance position by excluding utterance-initial and utterance-final items, rather than with base models including utterance position variables, for reasons explained below. The effects of intonational accent, which could not be controlled, are assessed by comparing results from the accent-coded subsample described above with the results for the effects of disfluencies and of predictability; see note 8 in Sec. IV A and note 11 in Sec. V A 1. Similarly, rather than controlling for the differences among the function words, we summarize the results of analyses for each of the words individually in the following sections, and discuss the behaviors of selected words in more detail.

IV. PLANNING PROBLEMS AND DISFLUENCIES

The production of speech is accompanied by a variety of disfluencies, whose characteristics have been extensively documented (Shriberg, 1994, *inter alia*). In particular, it appears that certain disfluencies often have a prospective source, occurring as a reaction to speakers' trouble in formulating an upcoming 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 neighboring words 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 problems in production. Fox Tree and Clark suggested that this relationship might extend to other words. Other work has also pointed to form effects in disfluent contexts. O'Shaughnessy (1992), for example, argued that words lengthen before pauses, and Shriberg (1995) showed that

forms of *I* and *the* were longer when they were repeated. It thus seems worthwhile to adopt the working hypothesis that longer and fuller forms are generally associated with planning problems, whether they function as signals of planning problems, or are part of production mechanisms to gain time to resolve planning problems, or some combination of the two.

In this section we extend such investigations to study the general relationship between disfluencies and pronunciation reduction in frequent words. Like Fox Tree and Clark (1997), we treat silent pauses, filled pauses *uh* and *um*, and repetitions as likely to be symptoms of planning problems. Each of the functors in our corpus is coded as belonging to a disfluent context if it is preceded or followed by one of these disfluencies.[5]

The following examples from our corpus illustrate the different disfluency contexts; numbers in parentheses are silence lengths in seconds.

Following disfluency	Sentence
Repetition	**I** I have strong objections to that.
Silence	...large numbers **of** (0.228) barefoot natives or something...
Filled pause (*uh*)	Somebody I talked to last week, they said they had **the** uh, they had problems doing some of the work

Preceding disfluency	Sentence
Repetition	I **I** have strong objections to that.
Silence	You know, the main things that I like about (0.214) **the** uh, job benefits...
Filled pause (*uh*)	it would encourage people, uh, **to** make more money

After eliminating uncodable items, there were 7999 function words coded for occurrence in preceding and following disfluent contexts. Of these, 2519, or 31 percent, occurred before or after a disfluency; 12 percent were followed by a disfluency, 16 percent were preceded by a disfluency, and four percent occurred between disfluencies.

A. Effects of disfluencies

Table IV compares durations, basic vowel frequency, reduced vowel frequency, and frequency of coda presence in fluent and disfluent contexts. Overall, longer and fuller forms are strongly associated with disfluencies, consistent with the hypothesis that they are symptoms of planning problems.

These observed differences, however, may not be a di-

TABLE V. Occurrence of disfluencies before and after function words. The percentages of following disfluencies are also shown.

		Disfluency before		
		Yes	No	Total
Disfluency after	Yes	297	927	1224
		24%	76%	100%
	No	1295	5480	6775
		19%	81%	100%
	Total	1592	6407	7999
		23%	77%	100%

rect indication of the effect of the disfluency since other factors affecting the form of words might be systematically associated with disfluencies. We therefore evaluated the effect of disfluencies in regression models after controlling for the control factors listed above in Sec. III F, for the predictability variables listed in Table IX below, and for relevant interactions among these variables.[6]

Neighboring disfluencies exert a strong influence on duration and on the frequency of full versus reduced vowels, in addition to effects of the control and predictability variables. They also moderately affect the frequency of basic vowels, but have no significant effect on coda deletion. The estimated magnitudes and significances of the effects are summarized as follows.

(i) **Duration:** words in disfluent contexts are 1.34 times longer [$F(1,6200) = 353.8$, $p < 0.0001$].

(ii) **Vowel reduction:** the odds of a word containing a full, unreduced vowel in a disfluent context are 1.68 times greater ($\chi^2 = 45.9$, $p < 0.0001$).

(iii) **Basic vowel:** the odds of a basic vowel form of a word occurring in a disfluent context are 1.23 times greater [$\chi^2(1) = 5.8$, $p < 0.02$].

Examining Table IV in more detail suggests that preceding and following disfluencies have different effects, and furthermore, that following disfluencies exert a stronger effect than preceding ones. We need to address the following questions.

(i) Is the effect of a disfluency before a word independent of the effect of one after the word?

(ii) When disfluencies occur before and after a word, are their effects cumulative? Multiplicative?

(iii) Are the effects of a disfluency after a word greater than the effects of one before a word?

Table V shows that disfluencies are more likely to occur in the presence of another disfluency. Although the increase

in the likelihood of a disfluency in one position given that one occurs in the other position is not large, the association is highly significant [$\chi^2(1) = 17.3$, $p < 0.0001$].

The effects are multiplicative. (Recall that the response variables in the regression analysis are logs of duration or of odds, so that additivity of factors in the regression model corresponds to multiplicativity of the untransformed variables.) Regressions with the variables for preceding and for following disfluencies show no significant effect for the interaction between the two. These results suggest, at least for duration and vowel reduction, that models with separate variables for preceding and following disfluencies are preferred to ones with a single variable for disfluencies in either position. The estimated magnitudes and significances of the effects are summarized in Table VI.

Since the effects are multiplicative, the effect on a word both preceded and followed by a disfluency is given by the product of effects in Table VI. For example, the estimated duration of such a word is 1.87 times that of a word not next to a disfluency (1.22×1.51, with rounding errors). Although the effects in Table VI are qualitatively comparable to those that could be derived from the uncontrolled observations in Table IV, they are in general smaller, and in some cases, much smaller. The estimated duration effects in Table VI, for example, 1.22 for preceding and 1.51 for following disfluencies, compared to effects of 1.30 and 2.10, respectively, derived from the observed average durations in Table IV.

Turning now to following versus preceding disfluencies, the effect of a following disfluency is greater than for a preceding one for duration [$F(1,6200) = 55.7$, $p < 0.0001$].[7] The difference between the two, however, is not significant for vowel reduction. Nor is it significant for the basic vowel variable. In summary, then, effects on duration are clearly best modeled with separate factors for preceding and following disfluencies. For these data, simpler single-factor models are adequate to account for the effects on the presence of full vowels and of basic vowels.[8]

Disfluencies appear to affect duration more strongly than the other measures of reduction. This pattern is repeated for the other factors that are discussed in successive sections. One obvious reason for this might be that the duration of a word encompasses all lenition factors, whereas the categorical variables target more specific ones. This raises the question of the interdependence of the response variables. Here, we focus on one important aspect of this general issue: Are the effects on duration simply consequences of the shortening effects of vowel reduction, nonbasic vowels, and coda deletion? The answer is emphatically no. As one would expect, all the categorical variables, especially vowel reduc-

TABLE VI. Estimated magnitudes and significance of the effects of disfluencies before and after a target word. The magnitudes for duration are the regression estimates of how much longer words are in the disfluent context. For the full vowel variable, they are estimates of the increase in the odds of occurrence of a full vowel in a disfluent context, compared to a fluent one.

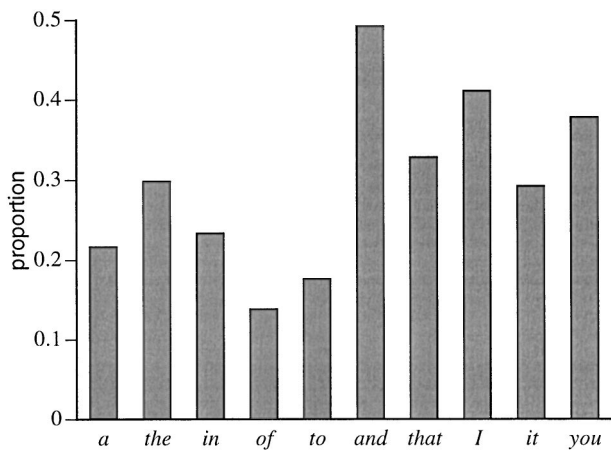|	Disfluency before		Disfluency after	
Response variable	Effect	Significance	Effect	Significance
Duration	1.22	$F(1,6200) = 120.5$, $p < 0.0001$	1.51	$F(1,6200) = 322.5$, $p < 0.0001$
Full vowel	1.59	$\chi^2(1) = 27.8$, $p < 0.0001$	1.68	$\chi^2(1) = 18.5$, $p < 0.0001$

FIG. 6. Proportion of occurrences in a preceding or following disfluent context for each function word. Overall, the proportion of words occurring in disfluent contexts is 0.32. The values are based on 8045 observations.

tion, do significantly affect duration. Nevertheless, after controlling for reduced and basic vowels, the effects of preceding and following disfluencies on duration are still very strong: 1.19 times longer after a disfluency [$F(1,6198)=101.3$, $p<0.0001$] and 1.48 times longer before a disfluency [$F(1,6198)=314.1$, $p<0.0001$]. Moreover, since there is no interaction between presence of preceding or following disfluencies and vowel reduction, disfluencies lengthen full vowels and reduced vowels in the same way.

There are a number of significant interactions of the disfluency variables with rate, context variables, age of speaker, and following word predictability variables. These interactions indicate that the effects of disfluencies vary to some degree for higher or lower values of the interacting variables. The effects are relatively small and mostly limited to effects on duration.

B. Items and disfluencies

The frequencies of occurrence of the ten function words in disfluent contexts vary widely. Figure 6 shows the proportion of observations of each function word in a disfluent context, either preceded or followed by a disfluency or both.[9] A general grouping by syntactic function is evident here. The complementizers/conjunctions *and* and *that* and the pronouns have the highest rates of occurrence in disfluent contexts, while the prepositions and the articles have the lowest rates. This suggests, not surprisingly, that syntactic class plays a role in the form and behavior of function words. Consideration of this issue is limited here to the remarks about some effects of the collocation *you know* and the binomial construction *X and Y* in Secs. IV B 2, V B 1, V B 2, and VI B below; see also Jurafsky *et al.* (2002).

More crucially for the assessment of an overall effect of disfluencies on reduction, the function words more likely to occur with disfluencies—*and*, *that*, *I*, *it*, and *you*—are in general both longer (especially *and* and *that*) and more frequent overall than the words occurring less frequently with disfluencies. This has the consequence that the average disfluent duration for all the words will be longer than the average fluent duration, even if there were no difference between each word's average duration in fluent and disfluent contexts. It is thus necessary to examine disfluency effects for the individual words before accepting the results of Sec. IV A above as valid.

Table VII summarizes the effects of disfluencies for the ten function words.

Examining first the effects on function word durations, longer durations are found in the presence of disfluencies for all ten of the function words, thus confirming the general effect. The effect of a following disfluency is more general than the effect of a preceding one, in parallel with the stronger overall effect found for following disfluencies. Since *in* is the least frequent of the words, failure to find a significant effect for a preceding disfluency is possibly due to the small sample; we did not explore other possibilities. There is, however, clearly no effect of a preceding disfluency for *you*. In Sec. IV B 2 below, it will be seen that this is likely due to two facts: (1) most of the preceding disfluencies occurred before *you know*, and (2) the *you* in *you know* is reduced rather than lengthened.

On the other hand, effects on vowel quality (whether the vowel was full or reduced, and whether full vowels were the word's basic vowel or another vowel) were spottier, judging from analyses of the individual words. The results support a general effect of less vowel reduction next to disfluencies,

TABLE VII. Significances of the effects of neighboring disfluencies on individual function words. Preceding and following disfluencies have been collapsed for the vowel reduction and basic vowel variables.

Effect on	a	the	in	of	to	and	that	I	it	you
Duration by a following disfluency	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Duration by a preceding disfluency	<0.0001	<0.0001	ns	<0.005	0.01	<0.0001	0.02	<0.0001	<0.0001	ns
Reduced vowel by any disfluency	<0.0001	<0.05	0.001	0.01	<0.02	<0.0001	ns	ns	ns	ns
Basic vowel by any disfluency	ns	0.02	ns	ns	ns	<0.01	ns	ns	<0.02	ns

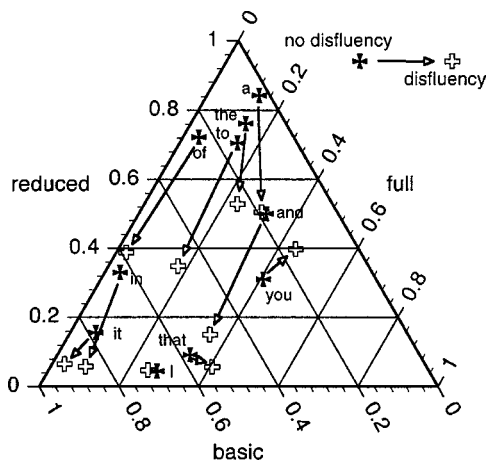Bell *et al.*: Disfluencies, predictability, position, and form variation

FIG. 7. Observed proportions of basic, full, and reduced vowels for the ten function words in nondisfluent contexts and in disfluent contexts. For each data point, the proportions of the three vowel categories sum to 1.0. Hence, the term full is used here in the special sense of not reduced and not basic. The proportions are based on 8045 observations, 5480 in nondisfluent contexts, 2565 in disfluent contexts.

since they reach significance for six of the function words for combined effects of preceding and following disfluencies, and, as we see below, the lack of significance for *I*, *it*, and *that* may be due to a ceiling effect. It is not clear that there is a general effect of more basic vowels next to disfluencies, though, since only three words show effects individually. The overall picture is easier to evaluate by combining the effects on reduced and basic vowels and examining them together for all the words, as presented in Fig. 7. (Note that in this figure, and in this section, we use the term ''full'' in the sense of nonbasic full, unlike the earlier use to mean any unreduced vowel.) In fluent contexts (indicated by filled crosses in the figure), the words vary greatly in the relative frequency of the vowel classes, basic, (nonbasic) full, and reduced, and this likely plays a role in how they are affected by neighboring disfluencies. In this figure, an arrow pointing down indicates that a word has fewer reduced vowels in disfluent contexts; if the arrow slants to the left, it also indicates that a word has a greater proportion of basic to nonbasic vowels in disfluent contexts. As one would expect from the overall results, most of the words exhibit one or both of these relations.

You is clearly anomalous, showing if anything an effect of disfluency in the opposite direction of the other words; some reasons for this surprising behavior are explored in Sec. IV B 2. The words along the left edge of the figure—*of*, *in*, and *it*—essentially have no nonbasic full vowels; hence, in a disfluent context, basic vowel frequency increases at the expense of reduced vowels. The lack of a significant increase in reduced vowel frequency for the three words at the bottom of the figure—*I*, *it*, and *that*—can possibly be attributed to their already very low rates of reduction, 16 percent or less. The lack of any increase in the basic vowels of *I* and *that*, on the other hand, seems to be a true item characteristic.

The four words at the top of the figure, which have the highest proportions of reduced vowels, over 70 percent, all showed significant decreases in reduced vowels, as might be expected. *The* is the only one of the four whose increase in

basic vowels is significant, although the observed leftward slants for *of* and *to* in the figure suggest stronger basic vowel effects for them. The large sample size for *the* may be a factor here. *A*, *of*, and *to* are among the less frequent of the function words and their many reduced vowels leaves few items to test basic vowel effects over ($n = 141$, 156, and 209, respectively). The significant basic vowel effect for *the* stands out in contrast, since its sample size is only modestly higher ($n = 268$).

And, which alone of the ten words has a relatively even balance among the three vowel categories, shows the strongest effects of disfluency contexts for both reduced vowels and basic vowels. Recall that it also has the highest rate of occurrence next to disfluencies.[10]

The overall picture suggests that there is generally less vowel reduction in the neighborhood of disfluencies, with the unexplained exception of *you*, possibly diminished in strength for items which already have few reduced vowels in fluent contexts. An increased number of basic vowels in disfluent contexts is clearly not general, and is likely to be a word-specific characteristic. Since contextual selection of lexical variants is an important source for variation between basic and other full vowels, further examination of how this differs for different words is warranted. Sorting out this and other differences will clearly take much more detailed study of the individual words and their contexts.

1. Initial disfluencies and *and*

Not only is *and* generally more frequent, longer, and more likely to occur with disfluencies, it is much more likely than the other words to occur in utterance-initial position, making up 48 percent of the function words there (Sec. VI A). This raises the question about the role of *and* in the preference, suggested by earlier research, for disfluencies to occur in initial positions.

Shriberg (1994), for example, showed that disfluencies were more likely to occur sentence initially than sentence medially, in three corpora (Switchboard, ATIS, and American Express) ($p < 0.0001$). In addition, Clark and Wasow (1998) suggested that disfluencies were more likely to occur at the beginning of large constituents like clauses than at the beginning of smaller constituents like words or phrases. This would presumably also result in a larger numbers of disfluencies in utterance-initial position. Results from our data on utterance-initial position agree with Shriberg's. After controlling for the variables mentioned in Sec. VI A below, we found that initial words in Switchboard are more likely to be disfluent than medial words ($p < 0.0001$). More specifically, filled and unfilled pauses (although not repetitions) are more likely to occur after the first word than after medial words ($p < 0.0001$). The greater likelihood of filled or unfilled pauses after initial words was, however, **only** true for *and*. For the other nine words, after removing the word *and*, there was no effect of increased disfluency rate on initial words. The rate of following disfluencies was the same for utterance-initial words and for noninitial words, 13 percent. Within our corpus the initial preference for disfluencies appears to be idiosyncratic to *and*. In a larger perspective it is likely to be related to the frequent use of *and* as a discourse

TABLE VIII. Observed average durations (ms) of function words in fluent and disfluent contexts. The number of observations appears in parentheses. The values are based on a sample excluding items beginning or ending a fragment, i.e., similar to the sample used in the regression analyses in this section.

	Another word		Silence		Filled pause		Repetition	
Preceded by	115	(5694)	145	(510)	147	(104)	201	(155)
Followed by	108	(5885)	187	(318)	307	(174)	186	(132)

marker. Discourse markers tend to occur initially in turns and utterances (Schiffrin, 1987). Perhaps such initial discourse markers tend to be followed by a filled or unfilled pause. A very preliminary survey over the entire 38 000-word set of Switchboard phonetic transcriptions supports this conjecture. First, turn-initial words are more likely to be followed by filled pauses or silence than noninitial words, 22 percent compared to 16 percent. Second, the vast majority of the initial disfluent words are words which frequently act as discourse markers. This suggests that the prevalence of silence and filled pauses in initial positions may be a fact more about discourse markers than about turn and utterance position.

2. The collocation you know

A number of characteristics of *you* stood out with respect to disfluencies: it was among the words most likely to occur with disfluencies, it was much more likely to occur after rather than before a disfluency than the other words, it showed no lengthening effect after a disfluency, and it showed no decrease of frequency of reduced vowels in disfluent contexts. All but the last of these can be attributed to the frequent occurrence of *you* in the collocation *you know*. This combination makes up 47 percent of the occurrences of *you* in our data. Since most of these are lexicalized fillers or editing terms, it is not surprising that the form of *you* tends to be reduced: *you* is about 25 percent shorter and about twice as likely to have a reduced vowel in *you know* than in other contexts.

You know itself very frequently occurs after a disfluency, which contributes to the apparent high rate of occurrence of *you* in disfluent contexts. Excluding *you know, you* is somewhat less likely than most of the function words to have a neighboring disfluency. The predominance of occurrence of *you* after rather than before disfluencies is partly an artifact of *you* in *you know* being almost always coded as having no following disfluency (pauses rarely separate the collocation), and partly because of the frequent occurrence of *you know* after a disfluency. In other contexts, *you* is only moderately more likely to occur after rather than before a disfluency.

The shorter and more reduced forms of *you* in *you know* obviously distorted the analyses of the effects of neighboring disfluencies. The reduced *you know*'s will count as fluent items for the following position, and hence will exaggerate the effect of a following disfluency. They will very frequently be among the disfluent items for the preceding position, and hence will dilute the effect of a previous disfluency. Indeed, when *you know* items are excluded, the effect of following disfluencies on duration is diminished, but it remains quite strong, especially considering the smaller sample $[F(1,296)=14.0, p=0.0002]$. And, without the *you know* items, a preceding disfluency appears to lengthen *you*

$[F(1,295)=4.8, p<0.05]$. One the other hand, it is still the case that *you* shows no decrease in the frequency of reduced vowels in disfluency contexts when the *you know* items are excluded. It is true that the overall rate of reduced vowels is decreased to 24 percent from 66 percent (cf. Fig. 5) by the exclusion of *you know*. Since this is still well above the levels of *I*, *it*, and *that*, it is not likely that a floor effect could keep the presence of a disfluency from reducing it further, as seems plausible for the lower reduced vowel rates of *I*, *it*, and *that*. *You*'s vowel reduction behavior thus remains an anomaly, all the more puzzling given the evident duration effects.

C. Differential effects of disfluency types

Does the effect of disfluent items on neighboring function words extend equally to each kind of disfluency that we have considered? We address this question here mainly to be assured that the effects described above are attributable in some degree to all of the disfluencies, in keeping with their assumed status as indicators of planning problems. The limitations of our database, which focuses on individual words in a very local context, precludes any analysis of the structure of disfluencies beyond the grossest details. One of the reasons for this is that disfluencies often are not simply silent pauses, filled pauses, or repetitions, but larger events combining some or all of these, as well as editing terms, as this Switchboard example shows:

> …built up in um PAUSE in the PAUSE in the
>
> PAUSE uh bureaucracy… .

Some of the more detailed questions about the form structure of disfluencies are treated in O'Shaughnessy (1992), Plauché and Shriberg (1999), and Shriberg (1994, 1999).

The observed average durations of function words in fluent and in different disfluent contexts are compared in Table VIII. The significances of duration differences reported below are, however, based on regression analyses controlled for the same variables described above. The durations for filled pauses and for repetitions in Table VII cover only the simple cases not combined with a silence.

When they precede a word, all three disfluency types have a lengthening effect. The significance of the effect is least for filled pauses $[F(1,5449)=11.0, p<0.001]$. The significances of the other effects are $p<0.0001$. The lengthening effect of a repetition is stronger than the effect of silences and filled pauses $[F(1,6058)=34.1, p<0.0001]$. The effects of silences and filled pauses do not differ significantly.

All three types also have a lengthening effect when they follow a word. Again, the effect is weakest, but nevertheless

Bell *et al.*: Disfluencies, predictability, position, and form variation

highly significant, for filled pauses [$F(1,5756) = 20.9$, $p < 0.0001$]. There are no significant differences between the effects of the different disfluencies, in spite of the apparently much longer durations before filled pauses.

D. Discussion

Function words which are preceded by or followed by disfluencies are longer and are more likely to have full vowels than words in fluent contexts. These effects are robust; all ten function words are longer when followed by disfluencies, and eight of ten when preceded by disfluencies. Effects on basic vowel frequency and coda presence, on the other hand, appear to depend on the lexical item or possibly, in the case of coda presence, the identity of the coda obstruent. Disfluencies after a word affect the word's form more strongly than disfluencies before a word. Preceding and following disfluencies tend to co-occur, and when they do, their effects are multiplicative. Finally, all three disfluency types have a lengthening effect.

V. WORD PREDICTABILITY FROM NEIGHBORING WORDS

In earlier work (Jurafsky *et al.*, 2001; Gregory *et al.*, 1999) we proposed the *probabilistic reduction hypothesis*: words are more reduced when they are more predictable or probable. In this section we focus on the extent to which the probability of a word given neighboring words affects reduction. There are many ways to measure the probability of a word. The simplest measure, *prior probability*, can be estimated from the relative frequency of the word in a sufficiently large corpus. The fact that the 10 words in this dataset were all very frequent, however, limited our ability to study relative frequency. The 3-to-1 range of frequency of the words is very small compared to the overall ratio of probability of about 100 000 to 1 for the highest and lowest frequency words in the entire 38 000-word phonetically transcribed portion of Switchboard. What variation there is, moreover, is inextricably confounded with the effects of form and patterns of combination of the individual items. Consequently, one cannot make useful inferences about the effects of relative frequency with the function words dataset.

We therefore limit our focus to the effect of neighboring words on predictability. Consider first the predictability of a word given the previous word. We use two measures of this. One is the *joint probability* of the two words $P(w_{i-1}w_i)$. The joint probability may be thought of as the prior probability of the two words taken together, and is estimated from the relative frequency of the two words together in a corpus. This is computed by counting the number of times the two words occur together, $C(w_{i-1}w_i)$, and dividing by N, the number of words in the corpus

$$P(w_{i-1}w_i) = \frac{C(w_{i-1}w_i)}{N}. \tag{1}$$

This is a variant of what Krug (1998) called the *string frequency* of the two words.

Used alone, joint probability is not an entirely satisfactory measure of word predictability. Pairs of words can have a high joint probability merely because the individual words are of high frequency (e.g., *of the*). But a word can occur infrequently, yet be very predictable every time it occurs. Thus most measures of predictability are based on metrics like *conditional probability* or *mutual information* which control for the frequencies of one or both of the words (Manning and Schütze, 1999). The second metric we use in this paper is such a metric: the *conditional probability of a word given the previous word*. This is also sometimes called the *transitional probability* (Saffran *et al.*, 1996b; Bush, 1999). The conditional probability of a particular target word w_i given a previous word w_{i-1} is estimated by counting the number of times the two words occur together $C(w_{i-1}w_i)$, and dividing by $C(w_{i-1})$, the occurrences of the first word

$$P(w_i|w_{i-1}) = \frac{C(w_{i-1}w_i)}{C(w_{i-1})}. \tag{2}$$

In addition to considering the preceding word, the effect of the following word may be measured by the two corresponding probabilities. The *joint probability of a word with the next word* $p(w_iw_{i+1})$ is estimated from the relative frequency of the two words together

$$P(w_iw_{i+1}) = \frac{C(w_iw_{i+1})}{N}. \tag{3}$$

Similarly, the *conditional probability of the target word given the next word* $p(w_i|w_{i+1})$ is the probability of the target word w_i given the next word w_{i+1}. This measures the predictability of a word given the next word the speaker is about to say, and is estimated by

$$P(w_i|w_{i+1}) = \frac{C(w_iw_{i+1})}{C(w_{i+1})}. \tag{4}$$

As we see below, while conditional probabilities are the most consistent of these factors affecting reduction, joint probabilities and the relative frequencies of surrounding words also contribute additional effects. It is thus helpful to consider their relationship with the conditional probabilities. The fundamental relationship among them is given by

$$P(w_i|w_x) = \frac{P(w_iw_x)}{P(w_x)}, \tag{5}$$

where w_x denotes either the preceding or the following word. (This can be derived from the definitions above.) Since we use log probabilities as factors in the regressions to assess effects, conditional probability as a single factor with weight B ($B \times \log$ conditional probability) is the same as the combination ($B \times \log$ joint probability) $-$ ($B \times \log$ relative frequency of the neighboring word). We can thus think about the conditional probability as combining the effects of joint probability and the relative frequency of the neighboring word under the simple assumption that they have equal (but opposite) weights. If we find that either joint probability or neighboring relative frequency [but not both; by Eq. (5) any third term of the three is redundant] contributes an additional effect, this tells us that the assumption of equal weights is incorrect, and that the combined effect of the probabilities is more complex. Since any two of the three probabilities in (5) capture all of the predictability effects of a neighboring

TABLE IX. Summary of probabilistic measures and high probability examples.

Measure	Definition	Examples
Joint of target with next word	$p(w_i w_{i+1})$	**you** *know*, **I** *think*
Joint of target with previous	$p(w_{i-1} w_i)$	*and* **I**, *in* **the**
Conditional of target given previous	$p(w_i \| w_{i-1})$	*rid* **of**, *kind* **of**
Conditional of target given next	$p(w_i \| w_{i+1})$	**I** *do*, **you** *know*
Conditional of target given surrounding	$p(w_i \| w_{i-1} \cdots w_{i+1})$	*matter* **of** *fact*

word, we have somewhat arbitrarily chosen to examine the conditional probabilities and the joint probabilities. Where both probabilities significantly affect reduction, we interpret the joint probability effect as either an indication that the joint probability is more heavily weighted than the neighboring word's relative frequency (in the case of less reduction with higher joint probabilities) or an indication that it is the relative frequency that is to be more heavily weighted (when there is more reduction with higher joint probabilities).

Table IX contains a summary of the probabilistic measures and some examples of high probability items from the dataset for each measure.

Other more complex conditional probabilities, often called *trigram probability* measures, played a smaller role in the analysis. Two of these were the *conditional probability of the target given the two previous words* $p(w_i|w_{i-2}w_{i-1})$, and the *conditional probability of the target given the two following words* $p(w_i|w_{i+1}w_{i+2})$. Neither of these turned out to have any effect on word forms. The other is the *conditional probability of the target given the two surrounding words* $p(w_i|w_{i-1}\cdots w_{i+1})$, estimated as follows:

$$P(w_i|w_{i-1}\cdots w_{i+1}) = \frac{C(w_{i-1}w_i w_{i+1})}{C(w_{i-1}\cdots w_{i+1})}. \tag{6}$$

We have also considered the *mutual information* (Fano, 1961) of the target word and the neighboring words in Gregory *et al.* (1999). There we showed that mutual information produces very similar results to the conditional probability of the target word given the neighboring word.

The actual computation for estimating these probabilities is somewhat more complex than the simple explanations above. Since the 38 000-word ICSI corpus is far too small to estimate word probabilities, they are estimated from the 2.4 million-word Switchboard corpus instead. We trained these probabilities via three separate stochastic grammars: a regular bigram grammar (conditioned on previous word), a reverse bigram grammar (conditioned on following word), and a centered trigram grammar. The counts were smoothed by Katz backoff with Good–Turing discounting (Jurafsky and Martin (2000), pp. 214–219, and references therein).

A. Effects of predictability

The main results that are reported in the following sections are based on regressions with the control variables listed above in Sec. III F, the preceding/following disfluency variables described in Sec. IV, and the relevant interactions among these variables. The reported results are based on a sample that excludes fragment-initial and fragment-final items, leaving a total sample of 6219 items. Separate analyses, which we do not report in detail, verified that adding additional control variables for the effects of prosodic position would not have materially changed the results. We thus chose not to use the smaller sample of some 4800 items of the ICSI corpus segmented by the LDC into sentence-like domains. See Secs. VI and VI A below for a fuller explanation of this sample and its prosodic coding. No analysis of predictability effects on basic vowel frequency or coda presence was attempted, partly because of the reduced sample sizes for those variables, and partly because they would reveal less about general effects of predictability, since their behavior varies much more from word to word.

1. Word duration

The predictability factors having the strongest effects on word duration are the conditional probabilities of the target word. As can be seen in Table X, both the conditional probability given the previous word $p(w_i|w_{i-1})$ and the conditional probability of the target word given the following word $p(w_i|w_{i+1})$ are highly significant factors. Target words which are more predictable are shorter. That is, the higher the conditional probability of the target given either of the neighboring words, the shorter the target word, as indicated by the effect magnitudes less than 1.0 in Table X. The shortening ratios used to measure effect magnitudes can be made more concrete by applying them to tokens which have typical values for other variables. This yields durations predicted by the regression models which include the other variables, as opposed to observed average durations, which are uncontrolled. Such words, if they are highly probable given the previous word (at the 95th percentile of the conditional probability), have a predicted duration of 90 ms; low conditional probability tokens (at the 5th percentile) have a predicted duration of 109 ms. The duration of words is affected similarly by their probability given the following word: highly probable tokens have a predicted duration of 86 ms; tokens with a low probability given the following word have a predicted duration of 116 ms.

There are also significant additional effects of the joint probabilities with previous and following words. When words have a higher joint probability with the following word, they are *longer*. This effect is in the opposite direction than the one we find with the conditional probabilities, and to some extent counterbalances the shortening effect of the conditional probability. In contrast, words with a higher joint probability with the previous word are shorter, affecting duration in the same way as the conditional probability. Moreover, there is a significant interaction between conditional probability given the previous word and joint probability with the previous word. This interaction captures some of the ways that the effects are uneven over the range of probabilities: joint probability has a shortening effect only for tokens whose conditional probability is above the median; and the shortening effect of conditional probability is greater for higher joint probabilities.

Thus predictability effects of the previous word and of the following word are similar in that both conditional prob-

TABLE X. Significances and magnitudes of effects of predictability variables on word duration and frequency of full vowels. The significance of each variable is obtained by adding it to a comparison regression model. The comparison model consists of the control variables for the preceding and following conditional probabilities; of the control variables plus the corresponding conditional probability for the joint probabilities; control variables plus the preceding conditional and joint probabilities for the interaction; and control variables plus all the other probability variables for the centered conditional probability. The values of F thus have degrees of freedom between $F(1,6197)$ and $F(1,6202)$. The effect magnitudes are ratios of length and ratios of odds of full vowels. They are estimated by evaluating the coefficients of the variables in the full regression equation over the range between the 5th and 95th percentiles of each variable. Effects for previous conditional and joint probabilities include the interaction, evaluated at median values of the variables.

Predictability variable	Duration			Full vowel proportion		
	Significance			Significance		
	F	p	Effect	$\chi^2(1)$	p	Effect
Conditional of target given previous	88.4	<0.0001	0.80	92.9	<0.0001	0.24
Joint of target with previous	43.7	<0.0001	0.94	55.2	<0.0001	2.44
Previous conditional×joint interaction	58.7	<0.0001		20.4	<0.0001	
Conditional of target given next	186.0	<0.0001	0.72	22.3	<0.0001	0.27
Joint of target with next	41.6	<0.0001	1.20	272.8	<0.0001	5.39
Conditional of target given surrounding	20.4	<0.0001	0.91	2.9	0.09	

abilities have shortening effects; they differ in that higher joint probability with the previous word shortens a word, but higher joint probability with following word lengthens it. In addition, no interaction was found between the previous word probabilities and following word probabilities.[11]

One further conditional probability affects word durations in addition to the variables above—the conditional probability of the word given both the previous and following words. Like the other conditional probabilities, tokens with higher conditional probabilities are shorter, but the effect is somewhat less. The predicted duration of tokens with high probabilities is 95 ms, whereas that of tokens with low probabilities is 104 ms. No significant contributions of probabilities were found involving the word before the previous word or the word following the following word (i.e., using the other trigram conditional probabilities described above). In other words, we are able to discern only strictly local probability effects, limited to the interaction of a word with the word next to it.

2. Vowel reduction

Neighboring word predictabilities affect vowel reduction in much the same way as they do word length. The conditional probability given the previous word and given the following word are both strongly associated with higher frequencies of reduction. The predicted likelihood of a full vowel in words which were highly predictable from the following word (at the 95th percentile of conditional probability) was 0.43, whereas the likelihood of a full vowel in low predictability words (at the 5th percentile) was 0.73. The predicted likelihoods for words with high and low predictability from the previous word were very similar, 0.43 and 0.72, respectively.

Again, there are also strong effects of the joint probabilities with previous and following words. For vowel reduction, however, a higher joint probability in *either* direction is associated with less reduction. Words with higher joint probabilities with either the previous or the following word are more likely to have full vowels, counterbalancing the reduc-

tion effect of the conditional probabilities. As with duration, the interaction between conditional probability given the previous word and joint probability with the previous word is highly significant, reflecting the same sort of variation in magnitude of effects that was described above for duration.

No significant additional effect was found from the preceding and the following words together. As with duration, there were no interaction effects between previous and following word predictability variables, nor were there any effects due to predictabilities involving words before the previous word or after the following word.

3. Interdependence of duration and vowel reduction

The strong effects of predictability on both shortening and on vowel reduction suggest that there may be separate sources for the two effects. Perhaps vowel reduction stems mainly from some sort of categorical choice in lexical production between full and reduced vowels, whereas shortening is mainly the result of gradient, noncategorical modifications at the level of phonetic encoding or of execution of the articulatory plans.[12] It is possible, however, that the shortening effects that we observe for function words might be solely a consequence of the vowel reduction effects, since reduced vowels are shorter than full vowels. If this were true, there might be no evidence for a gradient affect of probability on reduction. In order to test whether the effects of probability on shortening were completely due to vowel reduction, we added the full versus reduced vowel variable to the base model for duration as a control.

The probabilistic variables remain significant predictors of duration after controlling for vowel reduction. The vowel reduction variable of course accounts for a considerable amount of the duration variance (14.5 percent), so there should be less for the predictability variables to account for. Indeed, the predictability variables account for 3.8 percent of the variance in duration overall, but 2.6 percent of the variance in duration controlled for reduction. Nevertheless, except for the joint probability with the following word, all the individual predictability variables remain highly significant

TABLE XI. Significances of the effects of predictability variables on individual function words. Effects with significances above 0.01 are in boldface.

Effect on	a	the	in	of	to	and	that	I	it	you
Duration by conditional given following	**<0.05**	<0.001	<0.0001	ns	<0.0001	ns	<0.0001	ns	<0.005	<0.0001
Duration by conditional given previous	**0.05**	<0.001	ns	ns	0.0002	ns	ns	ns	**<0.02**	ns
Duration by joint with previous	ns	**0.02**	ns	<0.0001	<0.01	ns	ns	ns	**<0.05**	ns
Reduced vowel by conditional given following	ns	ns	0.0002	ns	0.0005	ns	ns	ns	ns	<0.0001
Reduced vowel by conditional given previous duration	ns	ns	**<0.05**	ns	ns	<0.0005	ns	**<0.05**	ns	ns

at levels of $p < 0.0001$. Predictability not only affects whether vowels are reduced or not, but it has an additional noncategorical effect on word duration.

Further confirmation results from an examination of the words with full and with reduced vowels separately, to see whether predictability shortening affects full vowels as well as reduced vowels. Even with the smaller subsamples, the probability variables remain highly significant at levels of $p < 0.0001$, with a few exceptions. The joint probability with the following word is a significant factor for reduced vowels ($p < 0.005$), but not for full vowels ($p = 0.15$); and conditional probability given the previous word is only marginally significant for full vowels ($p < 0.01$). We also verified that the possibly categorical deletion of final obstruents in the words *and*, *it*, *of*, and *that* did not account for the predictability effects on duration within the reduced and full vowel subsamples.

B. Variability of predictability effects by word

Individual analyses of the function words show that each word's duration is affected by one or more of the predictability variables. Table XI summarizes the effects on both duration and vowel reduction for the conditional probabilities given the previous word and given the following word. It also includes the effect on duration of the joint probability with the previous word. The most general effect is that of the conditional probability given the following word, affecting the duration of six of the words. The words showing no effect or only a marginal effect of this variable, *a*, *of*, *and*, and *I*, are scattered across functional categories and include both high- and low-frequency words. Thus it does not seem possible either to attribute the pattern of effects to limitations to particular classes of words or to attribute the exceptions generally to a lack of sensitivity of the analysis due to small sample sizes. The predictability variables involving the previous word clearly affect *the*, *of*, and *to*. In addition, the interaction between the conditional probability given the previous word and the joint probability with the previous word is a significant factor for five of the words. These include *of* and *to*, indicating that the conditional probability affects the duration of the word more when it occurs in a frequent combination with a previous word. The interaction is also a factor for *in* and *I*, suggesting that, although neither the conditional nor the joint probability is significant alone, that there may be an effect of the conditional probability for frequent combinations. Finally, there are marginal effects on duration of the bilateral conditional probability given previous and following words for *a*, *and*, *that*, and *to*, with significance values ranging from <0.01 to <0.05. As for vowel reduction

effects, it is evident that they are less general than those for duration. This parallels the pattern found for lengthening in disfluency contexts.

Overall, these results confirm the hypothesis that words in more predictable contexts have more reduced forms, since an effect for some predictability variable was found for each of the function words. On the other hand, the considerable variation in the strength of the effects (possibly none in some cases) underscores the importance of the interaction of each word's attributes with predictability. The hallmark of function words is that they are markers of particular pragmatic, semantic, and syntactic functions, and that they occur in particular classes of constructions. The kinds of constructions they occur in is bound to affect whether it is predominantly predictability from the left, from the right, or from both that they are subject to. Moreover, their occurrence in certain very frequent constructions may strongly influence the appearance of their overall sensitivity to predictability, since those constructions will necessarily be highly predictable contexts. While we do not explore these interesting connections here in detail, the discussions of high frequency uses of *and* and *you* that follow illustrate some of the interactions of a word's idiosyncratic behavior with predictability.

1. And *in binomial constructions*

One of the very frequent uses of *and* is as a conjunction to create binomial constructions such as *trucks and stuff*, *lockers and everything*. This immediately suggests a connection with the pattern of predictability effects on *and* discussed above, namely that *and* was one of few words to be affected by bilateral conditional probability (given both previous and following words). A very preliminary check confirms this. A fairly broad binomial category was coded by hand, which included modified and unmodified words, and adjectives and verbs as well as nouns. Excluding disfluent contexts, *and* is significantly shorter in binomials than elsewhere $[t(460) = 3.65, p < 0.0001]$. Furthermore, within binomials, *and* is significantly shorter when it is more predictable from the two surrounding words, whereas the bilateral conditional probability has no effect on the duration of *and* in its other occurrences.

2. You know *and predictability*

Recall from Sec. IV B 2 that 47 percent of the occurrences of *you* are in the collocation *you know*, and that in this context it is shorter and more likely to have a reduced vowel than in other contexts. *You* in *you know* is shorter (by about 25 ms) and much more likely to have a reduced vowel (50 percent compared to 24 percent) than *you* in other contexts.

The high frequency of the combination necessarily means that the predictability of *you* from following *know* is unusually high, 12.6 times other contexts. Its predictability from the preceding word, on the other hand, is lower, 0.42 times other contexts. This is presumably a consequence of fillers and editing terms occurring across a wide range of contexts, and hence being relatively unpredictable in any particular context. Recall from Table XI that *you* is strongly affected by predictability from the previous word, but little or not at all by the following word. The obvious question is whether this simply reflects the asymmetry of the *you know* combination, or whether it is more general. In contrast to the binomial *and* case, the results were little changed after excluding *you know: You* is shorter and more likely to have a reduced vowel when it is more predictable from the following word, but shows no effects of the predictability from the preceding word.

C. Discussion

Words that are more predictable are shorter and more likely to have reduced vowels, confirming the probabilistic reduction hypothesis introduced above. The conditional probability of the target word given the preceding word and given the following both play a role, in both duration and vowel reduction. The magnitudes of the duration effects are fairly substantial, in the order of 20 ms or more, or about 20 percent, over the range of the conditional probabilities (excluding the highest and lowest 5 percent of the items). The joint probabilities of the target words given the preceding and following words also played a role in reduction, as did the bilateral conditional probability of the target word given the two surrounding words. The local nature of the predictability variables is underscored by the lack of any effect involving words more than one word distant from the target word. The failure to find effects for all the probability variables on all the function words is possibly partly due to the smaller sample sizes, but the overall spotty pattern of effects indicates that there are real differences among the words. This sort of variation confirms the expectation that one source of the probability effects is the collocation of the function words in particular constructions. Are frequent collocations, perhaps semilexicalized, the only or primary source of the predictability observed here?

The answer seems to be no. In an earlier study (Jurafsky *et al.*, 2001), we showed that higher predictability is associated with increased reduction even in word combinations that are not lexicalized. We did this by looking at words with relatively low conditional probabilities, and showing that the effects of predictability from the preceding word hold not only for the more predictable cases, as would be expected if frequent collocations are the source of the effects, but also for the less predictable cases, which are unlikely to be lexicalized.

The fact that the effects of predictability on duration add to the effects on vowel reduction, and affect both full and reduced vowels, indicates that some of the effects of predictability on reduction are continuous and noncategorical. It is reasonable to conclude that predictability effects are not limited to lexical choice and combination at semantic and phonological form levels, so that the domains of applicability of the probabilistic reduction hypothesis include linguistic levels that allow continuous specification of phonetic form.

VI. THE POSITION OF A WORD IN PROSODIC DOMAINS

The location of a word in larger prosodic domains such as utterances, turns, intonational phrases, and phonological phrases plays an important role in reduction. Studies of language change and of pronunciation variation have long accepted three main effects—final lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*) initial strengthening (i.e., more extreme articulation) (Fougeron and Keating, 1997; Byrd *et al.*, 2000, *inter alia*), and final weakening (i.e., less extreme articulation) (Browman and Goldstein, 1992; Hock, 1986). During the last several decades more and more quantitative studies have helped make our understanding of these general effects more precise; see Fougeron and Keating (1997) for a review. Many of these results, however, derive from laboratory paradigms like reiterant speech, and have not been tested on natural speech production or over a wide range of lexical, prosodic, and pragmatic contexts. Furthermore, it has been difficult to tease apart prepausal lengthening from lengthening at the edge of prosodic domains.

To evaluate the effect that position in prosodic domains plays on function word reduction in conversational speech, as well as to control for positional effects in the analysis of other variables, we examine a word's position in an utterance-like domain. The domain we chose had already been transcribed for a large proportion of the Switchboard corpus by the Linguistic Data Consortium (LDC) (Meteer *et al.*, 1995), following the segmentation guidelines in Shriberg (1994). We use the term utterance for this LDC domain; Meteer *et al.* (1995) called them ''slash units.'' In general, these units are intended to model the sentence-like units which often make up spoken conversation, and hence are defined with respect to both syntactic coherence and an attempt at approximating large intonation boundaries. While this use of syntactic coherence as a heuristic for intonation boundaries is clearly inferior to a prosodic transcription of speech, the fact that grammatical boundaries and intonational boundaries are highly correlated (Croft, 1995) makes this methodological simplification less problematic.

The utterances include complete syntactic sentences.

(i)   I, I have strong objections to that.
(ii)   And that's not fair.
(iii)   Where, where are you?
(iv)   And, uh, I thought of those two things when I was, I was holding for a long time.

as well as phrases which function as complete turns

(i)   And, uh, until next time.
(ii)   A pop-up trailer, huh?
(iii)   The news.

In most cases an utterance was contained inside a single turn. Sometimes, however, an utterance was interrupted by a

TABLE XII. Duration and vowel reduction values for function words which are in initial position in the utterance, in final position, or in medial position (noninitial, non-final).

	Initial	Medial	Final
Duration (ms)	173	125	200
Vowel reduction	82.3%	57.4%	93.2%

backchannel such as *uh-huh*, or another remark from the interlocutor. In such cases, as in the following example, A's speech was counted as one utterance; thus, the word *and* is counted as utterance initial, but the word *here* is not.

> A: And, and I get mail
> B: Uh-huh.
> A: here at home under each of those names.

Larger turns are generally broken into utterances at syntactic boundaries which correlated with intonation boundaries.

> B: And, uh, I never really, messed with anything, uh, gardening or anything like that until now,
> B: but, uh, I, I keep hearing all the stories of, of different parts of town.

Readers interested in more details of the definition of utterances and the procedures followed by the LDC coders should see the coders' manual (Meteer *et al.*, 1995).

In general, utterance boundaries and turn boundaries were very highly correlated, as would be expected. For this reason, we did not examine turn-boundary position separately from utterance-boundary position. The edges of the LDC utterances should generally correspond with edges of intonational phrases (and also with edges of smaller units such as phonological phrases), whereas their interiors will sometimes contain words that are edges of intonational phrases as well as those of smaller units. Consequently, if utterance-edge strengthening effects are found, such results should be conservative.

A. Effect of utterance position

About two-thirds of the ICSI data had LDC utterance-boundary labels, so that 4777 observations were available for the analysis of utterance position.[13] Table XII shows observed values for duration and reduction in initial, medial, and final positions.

These observed differences, however, may not be valid indications of the effect of position in the prosodic domain, since other factors affecting the form of words might be systematically associated with prosodic positions. For example, Shriberg (1994) found that initial words are more likely to occur in the context of disfluencies. Since disfluencies cause words to be longer, this may exaggerate the actual effect of initial position. Initial position may have different kinds of segmental or accentual contexts than noninitial words, and may also be predictable in different ways. Pauses, which may be likely to occur after utterance-final position, would exaggerate the effect of final position. We therefore evaluated the effect of position in regression models after controlling for the factors listed in Secs. IV and V (and relevant interactions).

After controlling for all factors except predictability variables from the preceding word, initial words are longer than noninitial words [$F(1,4639)=30.1$, $p<0.0001$]. Initial words are also more likely to have a full (unreduced) vowel than noninitial words [$\chi^2(1)=192.9$, $p<0.0001$]. Conditional and joint probabilities with the preceding word were omitted from these analyses partly because they have no meaningful interpretation at the beginning of fragments, and if fragment-initial items were eliminated, the utterance-initial items would be halved, reducing the power of the analysis.

There is a more fundamental consideration, however. Low predictability is an expected characteristic of utterance-initial words, and can be expected to mask the effect of utterance position. This is the case. There remains no additional effect of initial position after adding the predictability variables as controls ($p=0.16$). (This analysis is based on the smaller subcorpus that excludes the fragment-initial items for which the predictability variables are not defined.) Low predictability, however, might well be considered to be an inherent characteristic of the position. This makes it unclear whether it is even appropriate to control for predictability. Analytically, the predictability variables mask the initial position effect, but the proper interpretation of this result awaits a deeper understanding of the interaction of predictability and prosodic domains than we possess.

Final position has long been known to play a role in lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*). As Table XII shows, the observed durations for final words are longer. After controlling for all factors except predictability variables from the following word, utterance-final words are longer than medial words [$F(1,3992)=25.5$, $p<0.0001$]. They are also more likely to have unreduced vowels [$\chi^2(1)=7.8$, $p=0.005$]. The utterance-final effect is not as sensitive to the masking from conditional and joint probabilities with the following word—utterance-final words are still longer [$F(1,3721)=12.7$, $p<0.0005$], and more likely to have unreduced vowels [$\chi^2(1)=7.7$, $p<0.01$] when controlled for these probabilities within fragments. Under those conditions, the estimated lengthening factor of final position is 1.23; and a word which occurred with a full vowel 60 percent of the time in medial position would have an estimated frequency of occurring with a full vowel in final position of 81 percent.

We do not report on the effect of position on the percentage of basic vowels or of coda deletion. Both these measures seem to be strongly affected by individual items. The results are difficult to interpret, but probably reflect specific high-frequency combinations of the function words with other words.

B. Variability of position effects by word

The final lengthening effect applies very generally; all ten function words are longer at the end of utterances. In contrast, only five words, *a*, *and*, *it*, *that*, and *the* have longer durations at the beginning of utterances than medially.

In addition, utterance-initial position is overwhelmingly dominated by the function words *and* and *I—and* makes up 48 percent, and *I* 32 percent of the function words in that position. Recall that *and* is also the longest of the function

words (Sec. III E). Is the combination of *and*'s length and frequent occurrence in initial position responsible for the utterance-initial lengthening effect above? Excluding *and*, an effect, although somewhat weaker, remains $[F(1,4078) = 6.8, p < 0.001]$. In addition, *and* alone shows a significant initial effect $[F(1,544) = 6.0, p < 0.02]$. On the other hand, there is no effect for *I* $[F(1,794) < 1]$. Thus, in contrast to our finding (Sec. IV B 1) that the association between initial position and disfluencies is limited to *and*, we conclude that the initial lengthening effect is not an artifact of the disproportionate number of longer *and*s initially, but applies more generally. The bias introduced by *and* simply exaggerates the general effect. The lack of an effect for *I*, however, indicates that there is no or little initial lengthening effect for some words, presumably due to idiosyncratic properties that we have not explored.

C. Discussion

The high-frequency function words studied here are longer and more likely to have full vowels at the beginning and end of the utterance-like domains coded by the LDC. Our results thus show that previous results on prosodic edge effects in laboratory speech (Fougeron and Keating, 1997, *inter alia*) can be extended to more natural conversational data. In addition, we found this lengthening after controlling for many contextual factors, including final pauses. This suggests that lengthening at prosodic edges plays a distinct role from prepausal lengthening. Initial strengthening is strongly associated with predictability from the previous word in ways whose understanding requires further research.

VII. CONCLUSIONS

Our results show that disfluencies, predictability, and utterance position all play strong and independent roles in whether a word is reduced, for all measures of reduction. While our regression study does not constitute a model in itself, these three results each have important implications for modeling of human lexical representation and production. First, a key result is that planning problems, as measured by disfluencies either preceding or following a function word, play a strong role in the word being longer and less reduced. This extends the results of Fox Tree and Clark (1997) on *the* to other function words. On the other hand, their suggestion that the basic form /ðiː/ may signal a disfluency appears to be lexically specific, since we found increases in basic vowel frequencies in disfluent contexts only for *the*, *and*, and *it*. More crucially, the influence of planning problems is extended to duration, a nonphonological measure of reduction, which appears to hold generally for all the words examined.

Second, the result that function words are reduced when they are highly probable given neighboring words lends evidence to probabilistic models of human language processing (Jurafsky, 1996; Saffran *et al.*, 1996a; Seidenberg and MacDonald, 1999). While some of this reduction may be due to lexicalization of multiword phrases, some of it is due to the mental representation of some kind of probabilistic links between words, since the effects are not limited to frequent collocations. Previous work has focused on the role of probability in comprehension. Our work shows how probability can play a related role in production.

Our results on probability also extend the work of Griffin and Bock (1998), who showed that interactions between predictability and frequency argue for what they called cascade theories of word production, and against discrete two-stage models of word production. In discrete two-stage models (Jescheniak and Levelt, 1994; Levelt *et al.*, 1999), the predictability of a word in context can help cause a word to be selected. But word selection is simply binary; once a word is selected, the amount of contextual predictability does not play a role in phonological encoding. By contrast, cascade theories (Dell, 1986; Stemberger, 1985) allow the amount of evidence causing a word to be selected to be passed to lower levels in word production. Our results show that highly predictable words are shorter even after controlling for reduction or deletion at the phonological level. This suggests that the extent to which the context predicts a word cannot just play a role at lexical selection or during the compilation of syntactic and prosodic frames. Predictability (and probably also some disfluency effects) must also make its way down to the level of articulatory routines.

Third, our results show that utterance-initial and utterance-final words are longer and less likely to be reduced than utterance-medial words. Since the effect of utterance position was significant even after controlling for pauses, our results show that final lengthening in conversational speech is an attribute of the prosodic or syntactic boundary condition itself, and not of the correlated presence of pauses at boundaries. On the other hand, while final lengthening is a separate effect from any lengthening from lower predictabilities in final position, a parallel separation of position and predictability for utterance-initial position was not found. This raises the question of the proper interpretation of the interaction of predictability from neighboring words and phrasal edges, that is, whether they should be considered separate but strongly associated sources of form variation, or whether the typically low predictability of words at phrasal edges should be regarded as an intrinsic attribute of the position.

Most contextual effects in speech, like assimilation, are strongest next to their source. The factors studied here are no exception, all being local in nature, involving the immediately previous or following word or an immediately previous or following utterance boundary. This is partly because the strategy of looking for effects in the most likely circumstances dictated that such contexts be examined first. Even so, there was no additional advantage of considering predictability from the previous pair of words instead of just the previous word, and similarly for predictability from following words. (And while we did not analyze utterance-second position, the observed average duration of words in second position did not differ from those in the other medial positions.) This does not mean that there are no effects of the sort considered here that are more global in nature. One example is the shortening of repeated words in a discourse reported by Fowler and Housum (1987), although this is unlikely to be an important factor for very high-frequency words, since

they are repetitions most of the time. But the strength of the local effects, together with the suggestions that there may be at least a sharp drop in the influence of more distant factors, indicates that the local-global dimension of effects deserves closer attention, both for its contribution to the structure of production models as well as its significance for speech processing applications.

Our results also have some implications for lexical representation, suggesting that multiple lexical representations of high-frequency function words may be more numerous than models of speech production have usually assumed. For example, in addition to the more commonly noticed allomorphy of *the* and *a*, allomorphic models should also be considered for at least *to*, *of*, and *and*. Furthermore, the selection of these variants 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).

Our results also have important implications for automatic speech recognition. Few of the factors that we show affect pronunciation variation are captured in current recognizers. Many of them could conceivably be added. Fosler-Lussier (1999a, 1999b) has shown first steps in this direction by showing how to build dynamic lexicons which are sensitive to speaking rate and the predictability of target words from previous words. These models could be extended to deal with predictability given following words. Similarly, planning problems could be handled with relatively simple modifications such as repetition detection and the use of a silence phone. The fact that there are key factors in reduction that are strictly local holds out the hope that good predictive models of word pronunciation may be based only on local information. We feel that these are promising directions for future investigations of ASR pronunciation models.

Much, of course, remains to be worked out in understanding the role of predictability in reduction. In addition to the exact locus of predictability in the cognitive processes involved in speech production, we still do not understand the complex interactions between conditional probabilities, joint probabilities, and item effects. Furthermore, we have simply reported first-order effects for probabilistic measures of local predictability, perhaps inviting the assumption that these effects are linear, holding in the same way from low to high probabilities. Even if this does not appear *a priori* unlikely to some, our own preliminary explorations of this question suggest that this simple model is not true. The more complex functional relationships between probability measures and reduction are yet to be determined.

If of course remains to be seen how general these effects are for all words in a conversation. In the general case, the relative frequency of each word, which we did not examine, plays a major role in the predictability of the word, and would be expected to influence word forms strongly. It may well interact with other measures, so that the effects found here might turn out not to be so strong for less frequent words. As a practical matter, disfluencies are disproportionately associated with function words, so that while we may find that longer words, less frequent words, and content words are longer and have less reduced forms in the presence of disfluencies, such occurrences may not be frequent enough to be of much practical importance for speech processing applications. Another difference that might be expected is that if the predictability effects found here are strongly associated with the connections of function words with particular constructions, then they may be weaker and less extensive for words that occur more freely.

In addition to these conclusions about lexical representation and production, we would like to end with a methodological insight. We hope to have shown that a corpus-based methodology such as ours can be paired with traditional controlled laboratory experiments to help provide insight into psychological processes like lexical production. Corpus-based methods have the advantage of ecological validity. The difficulty with corpus-based methods, of course, is that every possible confounding factor must be explicitly controlled in the statistical models. This requires time-consuming coding of data and extensive computational manipulations to make the data usable. Creating a very large hand-coded corpus is difficult, and there will always be factors that are beyond our ability to control for. But, to the extent that such control is possible, a corpus provides natural data whose frequencies and properties may be much closer to the natural task of language production than experimental materials can be. Obviously, it is important not to rely on any single method in studying human language; corpus-based study of lexical production is merely one tool in the psycholinguistic and phonetic arsenal, but one whose time, we feel, has come.

[1] The choice of citation vowel is clear, even across American dialects, for all the words except *of*, which likely varies idiolectally between [ʌ] and [ɑ]. The vowel [ʌ] is arbitrarily taken to be the basic vowel of *of* here.

[2] It is noteworthy that this does not accord with Bolinger's lengthening rule, which predicts that full vowels are longer before full vowels, whether separated by consonants or not. Testing the effect of following full vowels for just items with full vowels also shows no overall effect on duration.

[3] Since the regional dialect area of Switchboard speakers was coded in our database, we checked the effect of this variable on our reduction indicators. No effect of dialect was found for duration, vowel reduction, or coda deletion. Only the frequency of basic vowels appeared to differ across dialects. We did not pursue effects of other factors, individual comparisons of dialects, or item effects.

[4] It is perhaps surprising that much the same difference between men's and women's speech rate is found for both read speech (i.e., Byrd's TIMIT result that men spoke 6.2% faster) and conversation (men's rate of 5.4 syllables/s is 8.0 versus women here). This may be in part due to the local measure (i.e., between pauses) used here. It is more like an articulation rate

measure than longer-term speaking rate measures, which would be strongly influenced by pause rate.

[5]We followed Fox Tree and Clark (1997) in choosing this definition of disfluency mainly for simplicity; deciding if a word was preceded or followed by a disfluency could be coded automatically by software, and required no subjective coding. There are problems with this simplified definition. Obviously not all instances of these disfluencies reflect planning problems. Some pauses and repetitions are fluently planned, and filled or silent pauses may initiate repair of previous speech, to mention just two alternatives. In addition, this definition means that we did not code for other disfluencies such as cutoffs and restarts, or for editing phrases such as *I mean*. Incomplete and imprecise as our disfluency set is, it is nevertheless an index of aspects of conversational structure that are strongly linked to reduction variation in word forms.

[6]The control variables used in the regressions were actually a subset of these, since not all were significant factors for each one of the response variables. The regressions exclude items at the beginning or end of fragments and thus cover samples approximately 10 percent smaller than those in Table IV. We chose not to control for utterance position in the regressions used to estimate the effects of disfluencies in this section because of the smaller sample it would entail. We did, however, verify that for the smaller sample coded for utterance position, the effect of disfluencies is much the same with or without utterance position control. See a further discussion of utterance position in Sec. VI.

[7]Tests for the difference are based on the comparison of a regression model that includes both variables with one with a single variable summing the two, forcing equal weighting of the two variables.

[8]Similar results were found for the 385 unaccented words (Sec. III C). Function words which were in the context of disfluencies were longer than those which were not. There was a highly significant effect both for preceding disfluencies $[F(1,368)=10.4, p<0.002]$ and for following disfluencies $[F(1,369)=26.9, p<0.0001]$. This indicates, at least, that the main result cannot be an artifact of the distribution of intonational accents.

[9]Figure 6 does not differentiate preceding disfluencies from following ones, because for most of the words, one is about as frequent as the other. The exceptions are *of*, which is 2.9 times more likely to occur before a disfluency than after one, and *I* and *you*, which are, respectively, 2.4 times and 4.4 times more likely to occur after a disfluency than before.

[10]This summary, which for simplicity's sake has collapsed preceding and following disfluencies, conceals some strong differences among the words in the relative strength of effects, depending on the direction. For example, reduced vowels of *a* are strongly affected by both preceding and following disfluencies, but those of *and* are much more affected by preceding disfluencies, and those of *the* much more affected by following ones.

[11]As with disfluencies, similar results were found for the 385 unaccented words (Sec. III C). Duration was affected by the five predictability variables, i.e., joint and conditional probabilities with the preceding word and with the following word, and the interaction between the preceding joint and conditional probabilities. The overall effect was highly significant $[F(5,369)=4.4, p<0.001]$. Preceding and following word probabilities were also individually significant, and as with the overall sample, function words that were more predictable from neighboring words were shorter.

[12]Both vowel reduction and obstruent deletion can of course have gradient sources in speech production, and the transcribers' categorical choices for the variables are unable to distinguish whether the source is from a lexical choice or gradient articulatory variation. Particularly for some of the function words, some reduction and deletion is likely to be lexical, for example [tu] versus [tə], [æn] versus [ən], [əv] versus [ə], and forms of *and* with and without final [d].

[13]The utterance-segmented subset of Switchboard is comparable to the larger ICSI sample in terms of proportions of individual function words, average rate, average duration, and proportion of reduced vowels. It contains slightly more disfluencies, and many more men speakers, however.

Agresti, A. (**1996**). *An Introduction to Categorical Data Analysis* (Wiley, New York).

Bolinger, D. (**1986**). *Intonation and its Parts: Melody in Spoken English* (Stanford University Press, Stanford).

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

Bush, N. (**1999**). "The predictive value of transitional probability for word-boundary palatalization in English," Master's thesis, University of New Mexico, Albuquerque, NM.

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (**2000**). "Phrasal signatures in articulation," in *Papers in Laboratory Phonology V* (Cambridge University Press, Cambridge), pp. 70–87.

Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **23**, 39–54.

Clark, H. H., and Wasow, T. (**1998**). "Repeating words in spontaneous speech," Cogn. Psychol. **37**, 201–242.

Croft, W. (**1995**). "Intonation units and grammatical structure," Linguistics **33**, 839–882.

Crystal, T. H., and House, A. S. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.

Dell, G. S. (**1986**). "A spreading activation theory of retrieval in sentence production," Psychol. Rev. **93**, 283–321.

Fano, R. M. (**1961**). *Transmission of Information; A Statistical Theory of Communications* (MIT Press, Cambridge, MA).

Fidelholz, J. (**1975**). "Word frequency and vowel reduction in English," in *CLS-75* (University of Chicago, Chicago), pp. 200–213.

Fosler-Lussier, E. (**1999a**). "Contextual word and syllable pronunciation models," in Proceedings of the 1999 IEEE ASRU *Workshop*, Keystone, Colorado.

Fosler-Lussier, E. (**1999b**). "Dynamic Pronunciation Models for Automatic Speech Recognition," Ph.D. thesis, University of California, Berkeley. Reprinted as ICSI Technical Report TR-99-015.

Fosler-Lussier, E., and Morgan, N. (**1999**). "Effects of speaking rate and word frequency on conversational pronunciations," Speech Commun. **29**, 137–158.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

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

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

Godfrey, J., Holliman, E., and McDaniel, J. (**1992**). "SWITCHBOARD: Telephone speech corpus for research and development," in *Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92)* (IEEE, San Francisco), pp. 517–520.

Greenberg, S. (**1997**). "Switchboard transcription system," unpublished manuscript labelers' manual, revision of 19 February, 1997.

Greenberg, S., Ellis, D., and Hollenback, J. (**1996**). "Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus," in Proceedings of the International Conference on Spoken Language Processing (ICSLP-96), Philadelphia, PA, pp. S24–27.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (**1999**). "The effects of collocational strength and contextual predictability in lexical production," in *CLS-99* (University of Chicago, Chicago), pp. 151–166.

Griffin, Z. M., and Bock, K. (**1998**). "Constraint, word frequency, and the relationship between lexical processing levels in spoken word production," J. Memory Lang. **38**, 313–338.

Hock, H. H. (**1986**). *Principles of Historical Linguistics* (Mouton, The Hague).

Jescheniak, J. D., and Levelt, W. J. M. (**1994**). "Word frequency effects in speech production: Retrieval of syntactic information and of phonological form," J. Exp. Psychol. Learn Mem. Cogn **20**, 824–843.

Jespersen, O. (**1922**). *Language* (Holt, New York).

Jurafsky, D. (**1996**). "A probabilistic model of lexical and syntactic access and disambiguation," Cogn. Sci. **20**, 137–194.

Jurafsky, D., Bell, A., and Girand, C. (**2002**). "The role of the lemma in form variation," in *Papers in Laboratory Phonology 7*, edited by N. Warner and C. Gussenhoven (Mouton de Gruyter, Berlin/New York) pp. 3–34.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229–254.

Jurafsky, D., and Martin, J. H. (**2000**). *Speech and Language Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Keating, P. A., Byrd, D., Flemming, E., and Todaka, Y. (**1994**). "Phonetic analysis of word and segment variation using the TIMIT corpus of American English," Speech Commun. **14**, 131–142.

Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in a connected discourse," J. Phonetics **3**, 129–140.

Krug, M. (**1998**). "String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change," J. Engl. Linguistics **26**, 286–320.

Ladd, D. R., and Campbell, N. (**1991**). "Theories of prosodic structure: Evidence from syllable duration," in Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence, France, pp. 290–293.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (**1999**). "A theory of lexical access in speech production," Behav. Brain Sci. **22(1)**, 1–75.

Lieberman, P. (**1963**). "Some effects of the semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–175.

MacDonald, M. C. (**1993**). "The interaction of lexical and syntactic ambiguity," J. Memory Lang. **32**, 692–715.

Manning, C. D., and Schütze, H. (**1999**). *Foundations of Statistical Natural Language Processing* (MIT Press, Cambridge, MA).

Marcus, M. P., Santorini, B., Marcinkiewicz, M. A., and Taylor, A. (**1999**). *Treebank-3*, Linguistic Data Consortium (LDC). Catalog #LDC99T42.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (**1998**). "Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension," J. Memory Lang. **38**, 283–312.

Meteer, M. *et al.* (**1995**). *Dysfluency Annotation Stylebook for the Switchboard Corpus*, Linguistic Data Consortium. Revised June 1995 by Ann Taylor. ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL_book.ps.gz

Neu, H. (**1980**). "Ranking of constraints on /t,d/ deletion in American English: A statistical analysis," in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York), pp. 37–54.

O'Shaughnessy, D. (**1992**). "Automatic recognition of hesitations in spontaneous speech," in Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92), Vol. I, pp. 593–596.

Plauché, M., and Shriberg, E. (**1999**). "Data-driven subclassification of disfluent repetitions based on prosodic features," in Proceedings of the International Congress of Phonetic Sciences (ICPhS-99), San Francisco, Vol. 2, pp. 1513–1516.

Rhodes, R. A. (**1992**). "Flapping in American English," in *Proceedings of the 7th International Phonology Meeting*, edited by W. U. Dressler, M. Prinzhorn, and J. Rennison (Rosenberg and Sellier, Turin), pp. 217–232.

Rhodes, R. A. (**1996**). "English reduced vowels and the nature of natural processes," in *Natural Phonology: The State of the Art*, edited by B. Hurch and R. A. Rhodes (Mouton de Gruyter, The Hague), pp. 239–259.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996a**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996b**). "Statistical cues in language acquisition: Word segmentation by infants," in COGSCI-96, pp. 376–380.

Schiffrin, D. (**1987**). *Discourse Markers* (Cambridge University Press, Cambridge).

Schuchardt, H. (**1985**). *Über die Lautgesetze: Gegen die Junggrammatiker*. Robert Oppenheim, Berlin. Excerpted with English translation in *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change*, edited by T. Vennemann and T. H. Wilbur (Athenaum, Frankfurt, 1972) pp. 39–72.

Seidenberg, M. S., and MacDonald, M. C. (**1999**). "A probabilistic constraints approach to language acquisition and processing," Cogn. Sci. **23**, 569–588.

Shattuck-Hufnagel, S., and Ostendorf, M. (**1999**). POSH labeling guide—version 1.0. Unpublished draft.

Shriberg, E. (**1994**). "Preliminaries to a Theory of Speech Disfluencies," Ph.D. thesis, University of California, Berkeley, CA. (unpublished).

Shriberg, E. (**1995**). "Acoustic properties of disfluent repetitions," in Proceedings of the International Congress of Phonetic Sciences (ICPhS-95), Stockholm, Sweden, Vol. 4, pp. 384–387.

Shriberg, E. (**1999**). "Phonetic consequences of speech disfluency," in Proceedings of the International Congress of Phonetic Sciences (ICPhS-99), San Francisco, Vol. I, pp. 619–622.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (**1992**). "TOBI: a standard for labelling English prosody," in Proceedings of the International Conference on Spoken Language Processing (ICSLP-92), Vol. 2, pp. 867–870.

Stemberger, J. (**1985**). "An interactive activation model of language production," in *Progress in the Psychology of Language*, edited by A. Ellis (Erlbaum, London), pp. 143–186.

Trueswell, J. C., and Tanenhaus, M. K. (**1994**). "Toward a lexicalist framework for constraint-based syntactic ambiguity resolution," in *Perspectives on Sentence Processing*, edited by C. Clifton, Jr., L. Frazier, and K. Rayner (Erlbaum, Hillsdale, NJ), pp. 155–179.

Wald, B., and Shopen, T. (**1981**). "A researcher's guide to the sociolinguistic variable (ING)," in *Style and Variables in English*, edited by T. Shopen and J. M. Williams (Winthrop, Cambridge, MA), pp. 219–249.

Zipf, G. K. (**1929**). "Relative frequency as a determinant of phonetic change," Harv. Studies Classi Philol. **15**, 1–95.

Bell *et al.*: Disfluencies, predictability, position, and form variation