to appear in Journal of Memory and Language.

Predictability Effects on Durations of Content and Function Words in Conversational English

Alan Bell, Dept. of Linguistics and Inst. of Cognitive Science, University of Colorado, Boulder
Jason Brenier, Cataphora Inc., Redwood City, California
Michelle Gregory, Pacific Northwest National Laboratory, Richland, Washington
Cynthia Girand, Silver Creek Systems, Westminster, Colorado
Dan Jurafsky, Dept. of Linguistics, Stanford University

Abstract

In a regression study of conversational speech, we show that frequency, contextual predictability and repetition have separate contributions to word duration, despite their substantial correlations. Moreover, content- and function-word durations are affected differently by their frequency and predictability. Content words are shorter when more frequent, and shorter when repeated, while function words are not so affected. Function words have shorter pronunciations, after controlling for frequency and predictability. While both content and function words are strongly affected by predictability from the word following them, sensitivity to predictability from the preceding word is largely limited to very frequent function words. The results support the view that content and function words are accessed differently in production. We suggest a lexical-access-based model of our results, in which frequency or repetition lead to shorter or longer word durations by causing faster or slower lexical access, mediated by a general mechanism that coordinates the pace of higher-level planning and the execution of the articulatory plan.

This research was partially supported by the NSF, via awards IIS-9733067, IIS-9978025, and IIS-0624345, and by the Edinburgh Stanford Link. Many thanks to Ayako Ikeno, Bill Raymond, Eric Fosler-Lussier, and Dan Gildea for their important contributions to the compilation of the database and to Mirjam Ernestus, Susanne Gahl, Florian Jaeger, Rebecca Scarborough, and reviewers for many helpful suggestions. We are also very grateful to Stefanie Shattuck-Hufnagel and Mari Ostendorf for generously taking the time and effort to release to us a preliminary version of their prosodically coded portion of Switchboard.

1 Introduction

Two factors have long been assumed to play a crucial role in the way that words are produced and comprehended. One is a word's frequency or predictability. Another is the lexical category distinction between function words (closed-class words that play a grammatical role) and content words (open-class, semantically richer words). These two factors are intimately related, most obviously because function words are much more frequent and predictable than content words. This relation makes it difficult to know whether to attribute any particular effect in lexical production to words' frequency or predictability or to their lexical class. We propose to examine the two factors together, by analyzing how they affect the pronounced duration of words in natural conversational speech. The differences in the effects of frequency and predictability on content and function words, among other results, have consequences for how models of speech production treat these two factors.

The effects of frequency and predictability on lexical production are systematic, wide-ranging, and have been long observed. The reduced articulation of frequent words was pointed out by the Arab grammarians of the 12th century (Leslau, 1969) and was well-known by early modern linguists (Schuchardt, 1885; Jespersen, 1924/1965; Zipf, 1929; *inter alia*). In connected speech, frequent words have shorter durations and a variety of other lenited characteristics such as reduced vowels, deleted codas, more tapping and palatalization, and reduced pitch range (Fidelholz, 1975; Hooper, 1976; Rhodes, 1992, 1996; Bybee, 2000; Fosler-Lussier & Morgan, 1999; Pluymaekers, Ernestus, & Baayen, 2005b; Aylett & Turk, 2006; Munson, 2007; among others). For elicited speech in tasks such as picture naming, more frequent words have a shorter latency from picture display to the onset of the word, although no effects on duration have been found (Oldfield & Wingfield, 1965; Griffin & Bock, 1998; Damian, 2003; and references they cite). Shorter and more reduced responses for more frequent words have been found, in addition to shorter latencies in word-reading studies (Munson, 2007; Kello & Plaut, 2000, and references therein); Balota, Boland, and Shields (1989) found shorter durations for words preceded by related primes. More predictable words are also shorter and more reduced. This holds whatever predictability measure is used: from surrounding words in the sentence (Lieberman, 1963); cloze probability (Griffin & Bock, 1998; Liu et al., 1997); or the joint probability, conditional probability, or mutual information of a word with the previous or following words (Krug, 1998; Bybee & Scheibman, 1999; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999;

Jurafsky, Bell, Gregory, & Raymond, 2001; Aylett & Turk, 2004; Pluymaekers, Ernestus, & Baayen, 2005a). Longer-distance measures of the predictability of a word also affect surface form. For example, a **new** word/referent is less predictable when it occurs in a future sentence than an **old** or previously-mentioned referent. Indeed, English content words that have occurred in the previous context are shortened and/or less intelligible (Fowler & Housum, 1987; Fowler, 1988; Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Hawkins & Warren, 1994). Gahl and Garnsey (2004), showed grammatical effects of longer-distance predictability: in sentences with verbs that were strongly biased toward a particular syntactic subcategorization, the verbs and their arguments were both shorter when the subcategorization of the sentence matched the verb bias.

Less is known about differences in how content and function words are produced. The widely held presumption that they are produced differently rests mainly on evidence that they participate in different kinds of speech errors. As pointed out by Garrett (1975, 1980), word exchanges (as in the exchange of *trees* and *rain* in “Well, you can cut rain in the trees”) are very largely limited to content words. Function words, however, are common in shifts, a positional misplacement of a word (as in the shift of *out* from “figure out” in “If you can’t figure what that out is”). In addition, Garrett pointed out that sound exchanges (as the exchange of t and n in “pons and pats” for “pots and pans”, and of t and j in “Jom and Terry” for “Tom and Jerry”) are overwhelmingly restricted to content words. Processing of function words is impaired but content words are largely spared in agrammatic (Broca’s) aphasia, whereas the inability to process content words is characteristic of fluent (Wernicke’s) aphasics (Goodglass & Menn, 1985; Caramazza & Berndt, 1985). Such differences motivated Garrett’s proposal that function words have a privileged mode of access in speech production that is distinct from content words, an important issue that remains controversial.

There are also suggestions that there may be differences in the effects of frequency and predictability in the processing of content and function words, although only from studies of comprehension. Segalowitz and Lane (2000), examining word naming in sentence contexts, found a strong frequency effect for content words, possibly diminished for more frequent words. But there was a much weaker or no effect of frequency for function words, except possibly for the most infrequent ones. Segalowitz and Lane (2000) also found that function words were in general accessed faster than content words. This difference between content and function words was completely accounted for, however, by the frequency and cloze predictability of the target word being named. Herron and Bates (1997) found that contextual information facilitated recognition latencies for function words more than for content words. So frequency and predictability may affect access times of content and function words differently in comprehension, but some of the differences may be due to the different frequency and predictability profile of content and function words.

Thus despite this early experimental attention, important questions remain about how frequency and predictability, lexical class, and the relationship between them affect pronounced forms. In order to incorporate these factors into models of speech production, we need to know more about three key questions: (1) What are the effects of frequency, predictability, and repetition on word forms in speech production? (2) Do these effects differ for content and function words, and are there differences in how much reduction content and function words undergo, after controlling for predictability, frequency and related factors? (3) What implications do such effects of lexical class, frequency, predictability, and repetition have for production models?

We begin with question (1) on the effects of frequency, predictability, and repetition. The work described above has uncovered pervasive effects of reduction in frequent, predictable, and repeated words. Yet we still know very little about the relationship between these variables.

First, we know little about the relation between frequency and predictability. We don't even know whether there is a reduction effect of word frequency after controlling for word predictability, since previous work on predictability has not addressed crucial comparisons: Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, and Gildea (2003) investigated the effect of predictability on duration and reduction but not frequency. Pluymaekers et al. (2005a) used the predictability measure of mutual information, which confounds frequency and predictability. Aylett & Turk (2004) used word frequency and predictability from preceding words but did not report their individual effects for prosodically controlled data. Similarly, we know little about differences in the effects of frequency and predictability, both lexically and across the range of frequency and predictability. Griffin and Bock (1998), using picture naming in a sentential context, found that frequent words tended to be named faster, but not when they were highly constrained by context. The results of Gordon and Caramazza (1982) and Segalowitz and Lane (2000) suggest that, in comprehension at least, frequency effects may not be as strong for high-frequency words. Furthermore, there seem to be differences in the lexical generality of predictability effects. Some words were only affected by predictability from preceding contexts, and others only by predictability from following contexts, in the studies of 10 English function words (Bell et al., 2003) and of seven Dutch adjective/adverbs (Pluymaekers et al., 2005a).

Second, while we have known since the classic studies of Fowler and Housum (1987) and Fowler (1988) that a repeated word is (sometimes) reduced when compared with the first mention, we know very little about how this reduction is related to predictability and other variables. It is not clear whether the repetition effect holds after controlling for prosodic variables; previous studies attempting to answer this question have produced mixed results. Hawkins and Warren (1994), using words from conversations elicited by the use of objects and pictures, concluded that second-mention reduction could be accounted for by accent differences. Bard and Aylett (1999), however, did find both an accent effect and a repetition effect in map-task conversations, whereas Aylett and Turk (2004) found no repetition effect in addition to prosody effects, using the same sort of materials. Pluymaekers et al. (2005a) found shorter suffix durations of second and subsequent mentions for seven Dutch adjective/adverbs, after controlling for accent. Only the latter two studies have addressed the correlation of repeated words and their predictability. We also don't know whether the repetition effect is a binary one (given versus new), or whether it is scalar, with larger reductions for more repetitions. While some earlier studies such as Pluymaekers et al. (2005a) found an effect for the (log of) the number of prior mentions on duration, neither they nor earlier studies such as Gregory et al. (1999) and Aylett and Turk (2004) tested explicitly whether third and subsequent mentions contributed to it.

As for question (2), the systematic study of the relationship between lexical class and predictability variables in speech production has just begun. Frequency and repetition are known to affect reduction in content words, but their effect on function-word reduction is not known. Conversely, no study has considered whether function words undergo reduction effects more than content words, once prosody, frequency, and predictability are controlled. No studies have investigated whether word predictability affects content and function words identically. A similar situation exists for the reduction of repeated words. Studies of the effect have been mostly restricted to words used in referring expressions, mostly nouns, on the assumption that repetition reduction would parallel other processes that provide for the backgrounding of old information in discourse. The results of Pluymaekers et al. (2005a) for seven Dutch adjective/adverbs suggest that repetition reduction is not limited to words in referring phrases. But we still do not know whether verbs or adjectives in general would show a repetition effect, or what we would see with function words.

Once some basic empirical relations among frequency, predictability, lexical class, and reduction are established, we can address their implications for current models of speech production. There are two main issues here. The first is the role of lexical class: are content and function words accessed by the same mechanisms or by different ones? The second issue concerns the potential mechanisms or sources for the influence of words' predictability and frequency on their articulatory realizations. Unsurprisingly, the two issues are related.

The early and influential Garrett (1975, 1980) model proposed that content and function words have different modes of access. Function words are prespecified in syntactic templates. The syntactic templates are selected with forms of function words filled in, and content words are accessed at a later stage, filling lexical slots in the templates. As a consequence, content words, but not function words, could be subject to phonological exchanges, like the "pons and pats" example cited earlier, in accord with the observed absence of such errors in function words. This strong distinction has been challenged by Dell (1990). He concluded, from experiments with elicited speech errors, that both content and function words are subject to phonological errors. Such errors rarely occur in function words in naturalistic collections, however, because high-frequency words are generally less vulnerable to speech errors. The Extended Garrett Model of Lapointe and Dell (1989) maintains the differential access of content and function words in a weaker form. In their model, function words belong to syntactic fragments and are accessed via a feature-lookup procedure**.** Content words are accessed via network activation, filling slots in syntactic phrase structures. Evidence for these different mechanisms comes from observations of numerous characteristics of speech errors and from substitution patterns of function words in agrammatic aphasics. Consequently, at the level of lemma selection, content words are distinguished from function words. But at the level of phonological encoding, as segment sequences are specified, the two are no longer distinct, and thus have the potential of undergoing sound errors in the same way. Even this weaker distinction was rejected by Stemberger (1985), who proposed that content and function words are accessed in the same way, and that the speech error and aphasic patterns can be explained by differences in the function, form, frequency, and predictability of content and function words. It is plausible that differences in the frequency and predictability effects for content and function words would be relevant to this controversy. Indeed, we will argue that certain differences in these effects do support some degree of differential access.

The general issue of whether observed differences between content and function words are due to an intrinsic difference or to differences in their form and function has long vexed studies of the role of lexical class in speech processing. Stemberger's argument and Segalowitz and Lane's (2000) finding mentioned earlier that the faster access of function words in comprehension could be accounted for by their frequency and predictability are just two examples. Function words, in addition to marking syntactic and discourse structures, are more frequent, more predictable, have shorter phonological forms, are less likely to be prosodically prominent, and exhibit more idiosyncratic form variation than content words. See Goodglass and Menn (1985), Bock (1989), and Herron and Bates (1997) for further useful comments on this problem.

In examining mechanisms for probabilistic reduction, we adopt the widely accepted framework for models of speech production that comprises stages for conceptual and syntactic planning, lexical selection, phonological encoding and the construction of a prosodic frame, retrieval of articulatory routines, and articulatory execution. (Bock & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999; Dell & O'Seaghdha, 1992; Shattuck-Hufnagel, 1992; Wheeldon & Lahiri, 1997; Browman & Goldstein, 1990; among others). Within this framework three main classes of potential mechanisms or sources for the influence of words' predictability and frequency on their articulatory realizations have been proposed: the direct influence of frequency and predictability on articulation, the representation of word sequences at various stages, and lexical access.

Among direct-influence mechanisms, the automatization of articulatory sequences through repetition (Bybee & Hopper, 2001) is perhaps the simplest. Although it has plausible support from studies of highly practiced motor skills, it does not easily account for the range of effects of contextual predictability. Pluymaekers et al. (2005a) propose a second model, drawing from van Son, Koopmans-van Beinum, and Pols (1998) and Aylett and Turk (2004), in which the articulatory mechanism takes account of the frequency and predictability of words continuously, so that more redundant (less informative) parts of the speech stream are realized with less articulatory effort. A third possibility, the exemplar model of Pierrehumbert (2002), is more specific about the link from word to articulatory form. She proposes that a word's phonological encoding activates a region in the phonetic exemplar space, so that frequency effects would be directly represented by the distribution of a word's exemplars. Her model also includes a biasing mechanism for speaker, situation, and word effects that could be readily extended to accommodate contextual predictability effects.

The representation of predictable word sequences is a second candidate that could mediate the connection between frequency/predictability and articulatory form. Some predictable word sequences may be lexicalized or may be associated with routinized articulatory plans, leading to probabilistic reduction (Bush, 2001; MacWhinney, 2001). The fact that such multi-word representations are presumably limited to very frequent and/or very predictable word sequences leads one to expect stronger effects in the higher ranges of frequency/predictability, and a stronger effect of joint than of conditional probability. Grouping at the level of phonological encoding is another possibility that has not been considered. Predictable sequences may be more likely to be encoded within prosodic phrases and words rather than across them. If so, the positional effects on duration of prosodic phrasing would bias them toward greater reduction.

The third mechanism, lexical access, is a prime candidate for the source of frequency effects on form reduction, because of its known sensitivity to word frequency. Pierrehumbert (2002) and Levelt (2002), among others, have suggested a connection between the two, despite the fact that many lexical-access studies have found latency effects of frequency without concomitant duration effects (e.g., Damian, 2003).

But how would a higher level of activation of lexical forms, and thus faster access, lead to shorter pronunciations? This third mechanism would require some sort of feed-forward link from lexical access. The idea of such a link has been suggested by Pierrehumbert (2002) in the form of a parameter of "ease of retrieval" in phonological encoding and by Munson (2007) in relation to the longer and fuller articulation of disfluencies. One possibility drawing on these ideas is that short-term coordination between the access of words and their articulation may increase the strength and/or duration of articulation when the progress of phonological encoding is slowed, but not so impaired to trigger overt disfluencies. While a detailed examination of any such mechanisms is beyond the scope of this paper and must be the subject of future work, initial evidence that such a lexical-access mechanism is involved would be to establish that a word's frequency affects its articulatory form, even after controlling for predictability. We return to a discussion of all three mechanisms in the Discussion section.

Our goal in this paper, then, is to investigate the detailed effects of frequency and predictability on word forms, draw out their implications for the direct and indirect mechanisms for probabilistic reduction, and see whether they support differential access of content and function words. Natural conversational speech is an appropriate data source. It exhibits strong effects of frequency and predictability on articulatory forms, and equally importantly, the complex planning of utterances at higher levels is fully engaged with articulation. Our strategy is to use a single global measure of reduction, namely word duration, on the grounds that it is in general sensitive to most kinds of reduction (deletions, vowel reduction, reduced articulatory strength, etc.) whatever their locus in the word. On the other hand, we need to consider a fairly wide range of measures of frequency and predictability, given that they apparently affect words differently, and examine their effects over the full range of word frequencies. And, of course, content and function words need to be distinguished as well.

2 Materials and Method

The database for this study consists of 13,190 words from the segmentally transcribed ICSI subcorpus of the Switchboard corpus that have been transcribed for intonational accent and phrasing. Samples of one word from each intonational phrase are drawn from this 13190-word database, yielding datasets of about 950 to 1400 words depending on the class of words being considered. The effects of predictability factors on word duration were assessed by multiple regression analyses of word duration. Words in disfluent speech were excluded from analysis, as were words beginning or ending conversational turns. The factors of phonological form, articulation rate, intonational accent, intonational phrase position, and speaker age and sex were controlled by variables in the regression analyses. Predictability variables comprise word frequency; conditional and joint probabilities of a word with previous and following words; and repetition, whether or not a word has occurred previously in the discourse.

In the following sections, we first describe details of the database extracted from the Switchboard corpus in section 2.1. This includes an account of the exclusion of certain classes of words (2.1.1); the variables used to control the factors listed above (2.1.2); the computation of frequency, predictability, and repetition variables (2.1.3); and the coding of the lexical class variable (2.1.4). Section 2.2 describes the frequency and predictability characteristics of content and function words. The procedures used for the regression analysis of the data follow in section 2.3. The rationale for drawing samples from the database for the analyses and the procedures used are explained in section 2.4.

2.1 The database

The Switchboard corpus of 240 hours of conversational speech was collected at Texas Instruments in the early 1990s. It consists of telephone conversations between pairs of volunteer speakers on suggested topics, the text of which was transcribed at the word level (Godfrey, Holliman, & McDaniel, 1992). Approximately four hours of fragments (speech bounded by a turn and/or a 0.5 sec silence) were automatically extracted from the conversations and hand-transcribed at ICSI (the International Computer Science Institute, Berkeley), yielding a transcription of segments and syllables, with time markers for syllables, words, and pauses (Greenberg, Ellis, & Hollenback, 1996; Greenberg, 1997). Intonational accent and phrasing has been transcribed for about one-third of the ICSI corpus by Taylor, using the Tilt conventions (Taylor, 2000), and by Shattuck-Hufnagel and Ostendorf, using the Posh conventions (Shattuck-Hufnagel & Ostendorf, 1999). This subset of 13,190 words constitutes our database. It contains 1828 fragments from 924 conversations and 335 speakers. We coded the database for the part of speech of each word, social variables for each speaker, the presence of disfluent repetition strings, neighboring pauses and filled pauses, predictability variables, and a variety of phonetic and other control variables. Errors of transcription, alignment, and time marking of words were corrected or else the words were coded as ineligible. Words are defined by orthographic convention. Thus some frequent combinations of words, probably lexicalized as single compound words by many speakers (e.g. *peace corps, high school, ice cream*), are necessarily treated as separate words. Words with enclitics (e.g. *it's*, *you're*, *today'll*) are coded as single words, distinct from the unencliticized forms (*it*, *you*, *today*). Also transcribed as words in Switchboard are the filled pauses *uh* and *um* and the discourse markers *huh* and *uhhuh*. The database is thus a file of cases or items, each one an instance of a word coded for the analysis variables described in the following sections.

2.1.1 Exclusion of cases

Some of the items in the 13190-word database were incomparable in one way or another. These were marked as ineligible items, and were thus excluded from the datasets used for analyses. The filled pauses *uh* and *um* and the discourse markers *huh* and *uhhuh* were excluded. The discourse marker *yeah* was excluded; its very high frequency is largely due to its use in short back-channel utterances. The sequences *you know* and *I mean* were excluded because of their high frequency (137 and 20 instances in the 13190-word database) and usual use as discourse markers. Acronyms which were transcribed as one word for each letter, e.g. *T V*, *C N N*, *F M*, were excluded.

The largest group of exclusions was words in disfluent contexts. Words next to a pause, next to a filled pause *uh* or *um*, and words in disfluent repetition strings were excluded. A repetition string is a disfluent sequence that includes one or more repetitions of a word or word sequence, possibly with intercalated pauses, filled pauses, or editing terms. A simple example is *I I uh I*; more complex ones are *it's I mean you know it's* and *in um in the in the*. These were distinguished from planned repetitions such as emphatic repetitions like *good good* or sequences of certain constructions like *that that*, *is is*, *had had*. Words in disfluent contexts were excluded because the relation between duration and predictability is much more complex for them than for ones in fluent speech. In the presence of disfluencies, words are, on average, longer and less reduced; and function words, with high word frequencies, occur disproportionately often in these contexts (Fox Tree & Clark, 1997; O'Shaughnessy, 1992; Shriberg, 1995; Bell et al., 2003). And contextual predictabilities computed from adjacent words in disfluencies are not comparable to predictabilities in fluent speech.

Words beginning or ending an ICSI fragment (which included conversational turn beginnings and endings) were excluded from analysis, not only because of their special prosodic nature, but because there is no measure of their contextual predictability truly comparable to that for the fragment-internal words. Finally, short intonational phrases of three words or less were excluded, since a large proportion of them are back-channel utterances or other formulaic discourse responses (e.g. *oh good grief*). Overall, the items that were excluded comprised almost half of the items in the 13190-word database, leaving 6938 eligible items.

2.1.2 Control factors

Factors which were controlled by regression variables were word form, speech rate, intonational structure, and speaker characteristics. Variables for these factors were retained in the control regression model if their addition to the model improved its fit at a significance level of $p < .20$.

Word form is by far the most important source of variation in pronounced word duration. We considered three variables to account for word form: average word length, number of segments, and number of syllables. Average word length was estimated indirectly, by summing the average phone durations of the phones in the word's lexical transcription. The average phone durations were computed over a portion of the ICSI corpus of Switchboard that was segmented by phones, about 30 minutes and 45K phones long. Especially for infrequent words, this avoids the unreliability of a direct average over relatively few samples and contexts. A word's number of segments and number of syllables was taken from its dictionary transcription in the ICSI corpus and hence was invariant, not depending on the actual pronunciation. Various combinations of these three variables and their interactions can be used to control for word form. The simplest and most effective combination was log average word length (mirroring the use of log duration for the response variable) and number of syllables; number of segments was not a significant addition to these two ($p > .20$). The importance of number of syllables is consistent with many observations that syllable durations shrink as their number increases within a word, as in, for example, *luck* becoming successively shorter in *luck*, *lucky*, *luckily* (Lehiste, 1970).

Rate of speech was measured in (dictionary) syllables/sec, calculated within an ICSI fragment over pause-delimited stretches. It thus corresponds roughly to a measure of articulatory rate over a variable time domain. The actual variable used in the regression analyses was the transform log rate, since it was more effective than untransformed rate.

Binary contrast variables specified whether a word began an intonational phrase or not and whether it ended an intonational phrase or not. Initial phrase position was dropped from the analyses because it did not add significantly to the control model ($p > .20$). A single binary variable was used to code intonational accent, accented versus unaccented, ignoring various possible distinctions among types of accents. Each word was also marked according to the presence or absence of an accent on the preceding, prepreceding, following, and postfollowing words. These contextual accent variables had no significant effect and were not used in the analyses ($p > .20$).

The age and sex of the speaker also affected word durations. Older speakers have longer durations. Whether the speaker was a man or woman mainly influenced durations through a strong interaction with rate of speech, with men speaking at faster rates.

All the included variables except for speaker age ($p < .10$) were significant at the level of $p < .005$ or less; variables that were considered, but excluded, had significance levels of $p > .20$.

The control regression model thus consisted of eight variables:

- log average word length
- number of syllables
- log rate of speech
- intonational accent
- end of intonational phrase
- speaker age
- speaker sex
- speaker sex X log rate of speech.

2.1.3 Computing frequency, predictability, and repetition

The prior probability and predictability of words were estimated from counts of occurrences in the 2.4-million-word Switchboard corpus. Since these are counts of the written word forms they correspond more closely to counts of lexemes or forms than to lemma counts. Note, however, that they are not quite the same as counts of phonological forms, since they conflate the occurrences of homographs with different pronunciations (e.g. *lead* (N), *lead* (V)) and split the occurrence of homophones with different spellings (e.g. *sew*, *sow*). A word's probability of occurrence is estimated by its relative frequency, $C(w_i)/N$, the count of each word divided by the number of words in the corpus. The joint probability of a word w_i with the previous word w_{i-1} is estimated by the relative frequency of the two words together, $C(w_{i-1}w_i)/N$; the joint probability with the following word, by $C(w_iw_{i+1})/N$. The conditional probability of x given y is given by the ratio Prob(x and y)/Prob(y). Thus the conditional probability of a word given the previous word can be estimated by the ratio of two counts: $C(w_{i-1}w_i)$ divided by $C(w_{i-1})$. The conditional probability given the following word is estimated by $C(w_i w_{i+1})/C(w_{i+1})$. (In later discussion, we use the shorter terms "previous conditional probability" and "following conditional probability" to refer to the conditional probabilities of the current word given the previous word and given the following word, respectively.) Another measure of contextual probability, mutual information of x and y, is the ratio Prob(x and y)/Prob(x) Prob(y), and is estimated by the ratio of the appropriate counts.

Repetition of a word was represented by the number of times that the same lexeme had been previously spoken in the conversation, either by the speaker or by the listener. These were raw counts, including disfluent repetitions and homographs. Several transformations of the raw counts were used in the analyses. To avoid overweighting items with high counts, we used log counts and truncated counts (e.g., all values greater than or equal to 20 merged as the highest value). In order to assess whether more than one repetition had an additional effect, contrast variables coding various higher ranges of repetitions were combined with a basic binary contrast of first mention versus subsequent mention.

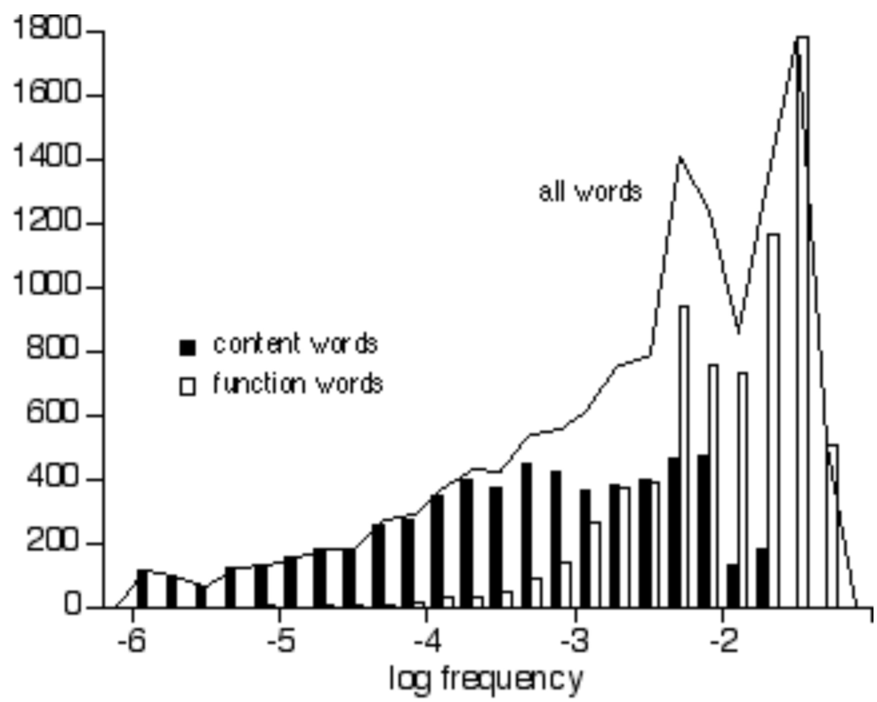
2.1.4 Coding the lexical class distinction

The distinction between content and function words was obtained by first using the Brill (1995) part-of-speech tagger to label all the words, and then editing the resulting word labels by hand. The corrections were fairly extensive and involved changes in appropriate contexts such as *okay* and *well* from modifier to discourse marker, *one's* from noun to pronoun, *that* from noun to demonstrative. We also examined words likely to occur in more than one category such as *so* (discourse marker, proform (as in *I thought so)*, or intensifier) and verb forms like *have*, *do*, *is*, etc. (verb or auxiliary), and corrected their labels. Then we collapsed the labelled categories into the binary distinction of content versus function. The function-word category included discourse markers, conjunctions, existential *there*, pronouns and other proforms, prepositions, articles, quantifiers, demonstratives, verb auxiliaries, and verb particles. The remainder, nouns, verbs, adverbs, and adjectives, made up the content-word category.

2.2 Frequency and predictability of content and function words

We use the logarithmic transform of frequency and predictability variables in our analyses. This is the usual practice for such variables, which range over several orders of magnitude and are typically asymmetrical and have long tails. In effect, we are considering relative differences in these variables, rather than absolute ones (e.g. treating the doubling of a frequency of .0001 to .0002 as equivalent to doubling .00001 to .00002, not a ten-times greater increase). Log word frequency in English is still asymmetrical, as well as bimodal, as can be seen from the histogram of word tokens in Figure 1. The figure also shows the distribution of function-word frequency and content-word frequency in the 13190-word database, and how much higher function-word frequencies are than content word frequencies. The double peak at the higher frequencies is

Figure 1. Histogram of the distribution of word tokens by log frequency. N = 13190.



largely a characteristic of function words, although content words also have a moderate second peak of higher-frequency words. The separate peak of highest-frequency words is composed of the 10 most frequent function words. Function words are not only more frequent, but more predictable than content words. This is confirmed in Table 1, a comparison of the median values of frequency and conditional probabilities of content and function words.

Table 1. Median $\log_{10}$ frequencies and conditional probabilities of content and function words.

factor	Content words	function words
word frequency	-3.37	-1.93
conditional prob. given previous	-2.41	-1.52
conditional prob. given following	-2.52	-1.38

The strong correlations between factors that affect word duration, namely lexical class, frequency and predictability, accent, and repetition, are displayed in Table 2. Because the

Table 2. Correlation coefficients for selected variables.

factor	content/ function	frequency	Previous conditional	following conditional	accent	repetition
duration	.61	-.70	-.53	-.59	.48	-.48
content/function		-.64	-.43	-.50	.41	-.43
word frequency			.69	.69	-.46	.62
prev. conditional				.45	-.36	.44
foll. conditional					-.37	.44
accent						-.30

measures of frequency and predictability are so highly correlated, it is hard to distinguish their effects. Although any single measure can be expected to show an effect on duration, such a finding alone cannot rule out that the effect is actually due to one or more of the others. The correlation of accent with the content/function distinction and with the predictability variables underscores the need to control for its effects on duration.

2.3 Regression analysis

Linear regression is the statistical tool employed to sort out how factors of interest influence the pronounced form of words. A regression model accounts for the variation in a response variable in terms of a linear combination of explanatory variables.

The response variable in this study is word duration. We chose this variable instead of other possible measures of lenition such as segment deletion or vowel reduction, both because it is sensitive to most kinds of form variation and because it is generally applicable to words of all types. Since we are normally interested in comparing relative differences in duration, rather than absolute ones (so that we do not, for example, equate a 30 ms change in a 50 ms word to a 30 ms change in a 300 ms word), the actual variable used in the regression analyses is the base-10 logarithm of word duration.

The explanatory variables can be divided for our purposes into two groups. The control variables, whose effects are known to be present and are not being investigated here, are the variables of word form, prosodic context, rate of speech and speaker characteristics described above in section 2.1. These variables account for 66 percent of the variation in log duration when content and function words are considered together.

The variables of interest are frequency, contextual-predictability variables, repetition, and the content-function distinction. Joint probability, conditional probability, and mutual information are closely related measures of contextual predictability. If a single measure of frequency and predictability is used, mutual information makes a good choice. To take the example of predictability from the previous word, recall that mutual information is the ratio of the frequency of a word occurring together with the previous word to the product of the individual frequencies of the word and of the previous word. It thus measures how much more or less likely it is that two words occur together than would be expected from just their overall frequencies. We did not use this measure, however, because its denominator contains the factor of word frequency, a factor of independent interest here. Instead, we used conditional probabilities and joint probabilities, which do not include the frequency of the target word in their definition. As a measure of contextual predictability, conditional probability is preferable to joint probability. Joint probability is the likelihood of two words occurring together; it has the disadvantage that it does not distinguish the combination of two unrelated high-frequency words from the combination of lower frequency words that commonly occur together. Thus we expected that conditional probability would be a more effective predictability factor. This turned out to be the case–joint probabilities with preceding and following words were not as effective predictors of duration as conditional probabilities, and did not make any significant addition to conditional probabilities in the regression models. They were thus dropped from the analyses and are not reported in the results.

The contributions of the explanatory factors can be ascertained by adding them to the control regression model. When, as is the case here, the explanatory variables are strongly correlated, not all of them can be expected to contribute significantly to the fit of the model. The factors which do make a contribution can be evaluated by comparing regression models of significant explanatory variables with and without the factor in question. The added factor is deemed to be a significant factor if improves the fit of the model sufficiently, as tested by the F statistic over the variance accounted for by the model. As indicators of the significance of a factor, we report the value of the F statistic, and its associated probability value. (For most results, only the approximate degrees of freedom of the F-statistic denominator are reported, not the exact value for each test, since it varies slightly, and in practice, inconsequentially, according to the number of variables in the model.) In addition to factor significance, we generally report ΔR^2, how much the addition of a factor increases R^2, the percentage of the variability accounted for by the regression variable. Since regression coefficients are notoriously unstable in models with highly correlated factors (Hocking, 2003, Ch. 5) we have not generally used them to report factors' estimated effects on duration.

2.4 Sampling and case selection

A fundamental assumption of multiple linear regression is that the cases analyzed are independent. This rules out analyzing all the word tokens of a conversational corpus because adjacent or neighboring words are surely mutually dependent in numerous ways. One way to avoid this sequential dependence is to restrict the analysis to a single word or small set of words. This is the approach used by Jurafsky et al. (2001) for content words ending in t or d, Bell et al. (2003) for 10 function words, and Pluymaekers et al. (2005a) for seven common Dutch words with the suffix *lijk*. The sets of words in the first two of these studies were large enough that dependence was only reduced, not avoided, since a fair number of the cases analyzed were in fact adjacent (e.g. *and it*, *child laughed*).

This study, which aims to assess predictability effects over broad classes of words, used another strategy, namely drawing samples of word tokens to avoid sequential dependencies. A number of sampling schemes are possible; for example, a simple one would be to choose every fourth (or fifth) word. We chose a more conservative procedure of taking only one word from each intonational phrase, reflecting the expected greater dependencies within phrases than across them, and avoiding possible positional biases. To do this, the 13190-word database was first divided into intonational phrases. (In technical terms, phrase breaks of level 3 were taken to delimit phrases.) Then the words in each phrase were marked as ineligible if they had been excluded for reasons discussed earlier in section 2.1. For each intonational phrase, one word was randomly selected from the eligible words in the phrase, if any. So, for example, separate samples were drawn for content words, for function words, and for all words without restriction to lexical class. We did not split the overall sample to obtain subsamples of content and function words, because this would have greatly reduced the power of the separate analyses. Conversely, pooling the content- and function-word samples to form an overall sample would have resulted in some words belonging to the same intonational phrase, defeating the original purpose of avoiding dependent cases. Thus the use of separate samples to investigate the two lexical classes (and different frequency ranges, as described below) maintains comparable power among the analyses and avoids sequential dependencies throughout.

Another violation of the case-independence assumption is that some of the word tokens in this data are spoken by the same speaker. Sampling one word from an intonational phrase reduces the number of word tokens per speaker considerably, but there are still clusters of word tokens for most speakers. For the sample of content words without high-frequency homonyms, there are 1238 word tokens and 282 speakers, about 4.5 word tokens per speaker. For most of the speakers, the clusters are small–more than three-quarters of the speakers have six or fewer word tokens in the sample. A few speakers have large clusters–five speakers have 18 or more tokens. The distribution of speakers and words in the other samples is similar. As a consequence, the significance probabilities reported in the results are inflated to some degree. This is unlikely to affect the main results, since they are very highly significant at probability levels of .0001 or less. Some of the results, however, are less secure. Nevertheless, as a conservative measure, we have retained factors with significances of $p < .05$ in our full models when assessing the significance of the additional contributions of the other factors.

Another potential problem is posed by the distribution of word frequency. Segalowicz and Lane (2000) point out that the log frequency distribution of their stimulus items (taken from a set of constructed sentences) is bimodal, with high frequency function words forming a second peak at the upper tail. The words in our conversational database are also bimodally distributed, as described above in section 2.2 and Figure 1. Moreover, the high-frequency words are exceptionally variable function words, and the control variable of word form, which is based on an assumed segmental representation, may not be as appropriate for them. To assure that the words in the upper frequency peak are not overly influencing patterns of probabilistic reduction, we examined the high-frequency function words that constitute the upper peak separately from the rest of the function words. These high-frequency words are the 10 most common function words whose form variation was examined by Bell et al. (2003) (*I*, *and*, *the*, *that*, *a*, *to*, *you*, *of*, *it*, *in*).

The upper frequency range of content words also contains some words which might skew the analysis, but for different reasons. The upper fifth frequency percentile of content words consists of *know, have*, and *so*. The unusually high frequency of *know* is largely due to its occurrence in the common discourse marker *you know*. The words of *have* and *so* have high frequencies because they are homonyms of very frequent function words (e.g. *have*, as main verb or auxiliary). The forms of *have*, *be*, and *do* are also common words which are content-function homonyms. Since they are frequent forms, with much of their frequency from occurrences as function words, they might have an anomalous effect on the analysis of content words. Accordingly, content words without these frequent content-function homonyms were analyzed separately to guard against this possibility.

For the analysis of content and function words combined, where content and function words are directly compared, the skewing of function words to the higher frequencies and of content words to the lower frequencies is the primary concern. To address this, the word-frequency range was partitioned into three: high-frequency words, mid-frequency words, and low-frequency words. An upper division point of $\log_{10}$ frequency = -2.1 (frequency = .008) was chosen at the 5th percentile point of content words in the 13190-word database and a lower division point of -3.1 (frequency = .0008) at the 95th percentile point of function words. This gives roughly equal frequency ranges of high, 34 percent; mid, 36 percent; and low, 30 percent of the words in the 13190-word database. Most (93 percent) of the high-frequency words are function words. The mid-frequency range is where function words and content words overlap; it is divided into 57 percent function words and 43 percent content words, very close to the 55/45-percent split for all frequencies. And the low-frequency words are mostly (90 percent) content words. Separate analyses were made for the full frequency range of words, for the mid/low-frequency words, and for the mid-frequency words.

There were between 4 and 14 cases with residual values greater than three standard deviations in the analyses. All cases were retained, since upon examination most appeared to be natural, just short pronunciations (e.g. *an* reduced to syllabic n in *on an old*, *think* as [ŋ] in *can't think of*) or long pronunciations (perhaps expressive). We chose not to second-guess the ICSI transcriptions of a few cases with ambiguous (e.g. vowel-vowel) syllable divisions.

In the course of this research we considered two alternative approaches worth mentioning. In addition to log frequency, we analyzed our data with the variable frequency rank, in case the nonnormal distribution of word frequency affected the regression analyses. The results were very similar to those reported for log frequency. In addition to the conditional probabilities given the previous word and given the following word, we also examined conditional probabilities given the previous two words, given the following two words, and given both the previous and following words. None of these were significant additions to the models ($F < 1$).

3 Results

The results for content words are presented first in section 3.1, based on a sample restricted to content words. The possible influence of certain high-frequency content words which have function-word homonyms (e.g. *have*, as main verb or auxiliary) is addressed here, as well as the influence of frequent collocations (e.g *kind of*). Results for function words follow in section 3.2, based on separate samples of function words. This section compares very-high-frequency function words with mid/low-frequency function words, and, again, treats the influence of certain frequent collocations. The two classes of words are then compared directly in section 3.3, based on results from a third sample of both content and function words. The analyses are based on separate samples drawn from the appropriate subsets of words in the entire 13190-word database. (See section 2.4.) Thus, for example, the content-word sample is not a subset of the sample of all words, and the sum of the content- and function-word sample sizes does not equal the size of the all-word sample. Results for all regression analyses are for models that include the control variables described in section 2.1.2.

3.1 Content words

For content words, following conditional probability and word frequency are the strongest factors affecting pronounced duration. There is also a strong interaction between word frequency and following conditional probability: the shortening effect of predictability from the following word is diminished for lower frequency words (and vice-versa). Previous conditional probability, however, is not a significant factor for content words. Repeated words have shorter pronunciations, but the difference is only nominally significant in this sample. Factor effects for the content-word sample are presented in Table 3.

Table 3. Predictability factors affecting duration of content words. Nonsignificant interactions are not included in the table.

factor	ΔR^2 (percent)	F(1, ~1275)	significance
main factors			
frequency	1.36	51.4	$p < .00005$
following conditional probability	1.13	42.7	$p < .00005$
previous conditional probability	.06	2.5	$p = .12$
repetition	.13	5.0	$p < .05$
interactions			
frequency X following conditional prob.	.50	19.1	$p < .00005$

In addition to the interaction of word frequency with following conditional probability, the quadratic variables of $(\text{word frequency})^2$ and $(\text{following conditional probability})^2$ are also significant additions to the model containing frequency, following conditional probability, and repetition. Regressions with any one of these interaction or quadratic variables indicate that the predictability and frequency effects on form reduction are stronger at higher levels of frequency and/or predictability. Only the interaction of frequency with following conditional probability has been retained in Table 3. It is the strongest of the three factors, and since they all compete with each other, a model with any two is no improvement. Nevertheless, the differences among these terms are small, so it is not possible to conclude for this data whether the increased effect is a combination of frequency and predictability, or just limited to frequency, or just limited to following predictability.

To assess the possible influence of the content words with frequent function-word homonyms, we analyzed a separate sample that excluded such items, as described above in section 2.4. Results from this dataset are shown in Table 4. The main effects for the probabilistic variables are essentially the same as those for the sample of all content words, confirming that they are not unduly influenced by the high-frequency homonyms. The factors that account for greater reduction at higher levels of frequency or predictability do differ, though. When the high-frequency content-function homonyms are excluded, $(\text{word frequency})^2$ is a highly significant factor, and is much stronger than either the interaction of frequency and following predictability or $(\text{following conditional probability})^2$, both $F < 1$. We conclude that the main source of the increased effects is word frequency, since it appears that effects involving following conditional probability were influenced by the inclusion of the high frequency content-function homonyms.

Table 4. Predictability factors affecting duration of content words excluding high-frequency homonyms.

factor	ΔR^2 (percent)	$F(1, \sim1220)$	significance
	main factors		
frequency	1.44	52.3	$p < .00005$
frequency^2	.68	25.2	$p < .00005$
following conditional probability	1.54	57.0	$p < .00005$
previous conditional probability	.01	< 1	NS
repetition	.24	8.7	$p < .005$

Without the high-frequency function-word homonyms, there is a stronger repetition effect for content words. The magnitude of the effect is relatively small, however–the estimated relative duration of subsequent mentions is 4.5% less than first mentions, accent and probabilistic variables being equal. The most effective repetition variable was the binary contrast between first mentions versus all subsequent mentions, which is reported in Table 4. Other measures of repetition were less effective: raw number of repetitions was not significant, $p = .10$. Measures which reduced the effect of multiple repetitions were better, the best of which was the log of the number of repetitions ($p < .01$). To address the question whether or not there is further shortening of mentions subsequent to the second one, a binary contrast variable opposing words which were second mentions to words which had occurred three or more times was added to the first-mention variable. (There are 678 first mentions in the sample; repeated mentions are split fairly evenly, 237 second mentions to 297 cases mentioned more than twice.) This variable made no significant addition to the fit of the model ($F < 1$). There is thus no evidence from this data that repetition reduction increases as the number of repetitions increases.

It might be expected that the effect would be found mainly with nouns, if it is one of repeated referents. There was no interaction of repetition with the contrast between nouns and verbs/adjectives/adverbs. Similar results came from separate analyses of the two classes (476 nouns, 736 verbs/adjectives/adverbs). Repeated nouns have shorter pronunciations, $F(1,466) = 4.1$, $p < .05$. Repeated verbs/adjectives/adverbs also have shorter pronunciations, $F(1,725) = 6.2$, $p = .01$. This data thus is consistent with the hypothesis that the repetition effect applies to the entire class of content words.

Another issue concerns the effect of the item of analysis, namely orthographic words. Many orthographic word sequences are surely single lexical entries. These include frequent collocations such as *lot of*, *kind of*, *in fact*, *of course*, as well as many compounds such as *income tax*, *New York*, *high school*, etc., which are conventionally written as separate words. Words in such sequences characteristically have very high conditional predictabilities, and presumably have shorter pronunciations than they would have in other contexts, by virtue of the weakening processes that are associated with compounds and similar constructions. If these collocations had been treated as single lexical entries, the durations would still be relatively short, but their word frequencies would have been much lower than the component words, and the high conditional probabilities between the components would not be present. The effect on the present analysis would be to decrease the shortening effects of both word frequency and conditional probabilities. (To some extent this would be compensated by an increase in the weight of the factor of syllable count.) It is thus important to establish that the frequency and predictability effects on duration are not artifacts of word definition. Reanalyzing the data with such collocations as single cases is not practical with this data. And while some collocations such as those mentioned can be identified with some assurance, there is such a gradation of collocation that one could not reasonably hope to identify exactly which ones are lexical entries. What is and isn't a lexical entry varies from one speaker to another; it is also plausible that many collocations can be retrieved from the lexicon either as word sequences or as single entries. We therefore used two modified analyses to ascertain the effect of collocations. First, from the sample of content words without high-frequency homonyms, we excluded 20 cases of *kind of* (in the sense of "somewhat") and *lot of*, since these were numerous and arguably single lexical words. Their exclusion changed the results of Table 4 only in minor details. Second, we excluded a larger number of cases with the highest values of mutual information, a bilateral predictability index that has been effective in identifying collocations (Manning & Schütze, 1999). (These were 195 cases out of 1238 whose values of the $\log_{10}$ of preceding or following mutual information was greater than 2.28.) The remaining cases contained very few sequences that we judged to be candidates for lexicalization. In this smaller sample, the effects of frequency and following conditional probability were a little weaker, as expected, but overall, the results were the same: Frequency, frequency^2, and following conditional probability were still highly significant ($p < .00005$); and repetition significance was still $p < .005$. This indicates that lexicalized collocations are but a minor contribution to frequency and predictability effects on content-word durations in this data; nor are they responsible for the greater degree of reduction at higher levels of frequency.

3.2 Function words

Recall that the high-frequency function words constitute a separate peak in the frequency distribution (section 2.2, Figure 1), and may have other special characteristics. We thus examine first a sample of function words that excludes the 10 highest-frequency function words, called the mid/low-frequency sample, since it is likely to best reflect the general characteristics of function words. Then we examine the high-frequency sample that consists of the 10 most frequent function words ($\log_{10}$ frequency < -2.0) listed earlier in section 2.2. An analysis of all the function words follows.

The relative durations of the 10 high-frequency words are sensitive to the segmental context of the following word, mainly because they are all monosyllables, half with open syllables, half with closed syllables. Five variables were added to the control model for this sample only to account for this influence: open/closed syllable, following consonant/vowel, following full/reduced vowel, and interactions of open/closed with the other two variables. (Following context is not as important, in addition to the other control variables, for the more complex and more variable word forms in the other samples.)

For the mid/low-frequency function words, following conditional probability is the only strongly significant factor. Previous conditional probability barely reaches .05 significance, and word frequency makes no contribution at all. The effects are summarized in Table 5.

Table 5. Predictability factors affecting duration of mid/low-frequency function words.

factor	ΔR^2 (percent)	F(1, ~935)	significance
main factors			
frequency	.00	< 1	NS
following conditional probability	1.23	17.7	p = .0001
previous conditional probability	.27	3.9	p < .05

For high-frequency function words, previous conditional probability is the most significant factor affecting word duration; word frequency is a moderately significant factor. See Table 6. The lack of any effect of following conditional probability is surprising. It is not

Table 6. Predictability factors affecting duration of high-frequency function words.

factor	ΔR^2 (percent)	F(1, ~945)	significance
main factors			
Frequency	.68	9.8	p < .002
following conditional probability	.00	< 1	NS
previous conditional probability	1.08	15.5	p = .0001
interactions			
frequency X previous conditional prob.	.53	7.6	p < .01

consistent with the significant following conditional probability effects found individually for seven of the same high-frequency function words in Bell et al. (2003). (Bell et al. also found an overall effect for following conditional probability. They did not control for intonational accent, however, nor were the cases sampled.) Interpreting the influence of word frequency in this sample is problematic. This is because with only 10 words, differences in their frequencies are necessarily confounded with individual differences in their phonological characteristics and the structures they occur in. Frequency for these words mainly distinguishes the three most frequent words, *I*, *and*, and *the*, from the rest. In particular, the control variable for word length, based on the average durations of the assumed underlying segments of each word, may not be so appropriate for this sample of highly variable words as it is generally. That this may be the case is supported by an alternative measure of word length, simply the average duration of each word over the ICSI corpus. Word frequency is not a significant factor for high-frequency function words in a regression with this variable ($F = 6.9$, $p = .09$).

The differences between the two frequency classes of function words are clarified by the analysis of a sample drawn from both classes, whose effects are summarized in Table 7. The models in this analysis include a binary variable (frequency split), which contrasts the high-frequency words with the remaining mid- and low-frequency ones. This variable is the strongest factor in the analysis, and displaces word frequency. (The frequency split factor was significant

Table 7. Predictability factors affecting duration of function words. Frequency split is the binary contrast between high frequency and mid/low frequency words.

factor	ΔR^2 (percent)	$F(1, \sim 1235)$	significance
	main factors		
frequency	.05	1.2	NS
frequency split	1.37	33.1	$p < .00005$
following conditional probability	.90	21.8	$p < .00005$
previous conditional probability	.80	19.5	$p < .00005$
	interactions		
frequency split X following cond. prob.	.44	10.8	$p = .001$
frequency split X previous cond. prob.	.17	4.2	$p < .05$
foll. cond. prob. X prev. cond. prob.	.18	4.5	$p < .05$

even in addition to a model including word frequency, whereas word frequency after frequency split was not. Nor was there any interaction between the two variables.) Both conditional probabilities are significant factors. The interactions, although at moderate significance levels, confirm that the effect of previous conditional probability is largely characteristic of the high-frequency function words, whereas the effect of following conditional probability is more associated with the mid- and low-frequency function words. Repetition of function words has no shortening effect, in contrast to content words.

Recall also that the effect of previous conditional probability did not occur with content words. This suggests that this effect arises to some extent from those function words that combine closely with preceding words. To explore this possibility, we looked at the high-frequency function words by their values of mutual information with the previous word, a measure which is useful for picking out collocations of words. It turned out that the word combinations with the highest values were dominated by *of* and *to*, and included numerous instances of special sequences such as *kind of*, *going to*, and *have to*, which are frequently encliticized, or have idiosyncratic pronunciations. Some, like those just cited, have noncomponential meanings. The high-frequency function words were reanalyzed in two ways to judge the effect of such cases. The first was to exclude forms likely to have idiosyncratic pronunciations: *kind of, sort of, lot of, some of, ought to, used to, going to, want to, have to,* and *got to*. Excluding 59 instances of these sequences reduced the significance of previous conditional probability ($\Delta R^2 = .47$, $F(1,886) = 6.5$, $p = .01$). The second analysis excluded 112 cases whose mutual information with the previous word was greater than 1.2. This eliminated most of the putative special sequences, but also sequences such as *live in, issued a, percent of,* and *hard to*, which, although highly predictable, do not seem to be candidates for special lexical forms, on the basis either of form or meaning. Previous conditional probability was not a significant factor in this analysis, ($\Delta R^2 = .20$, $F(1,831) = 2.5$, $p = .11$), presumably not just from loss of power, since the significance of word frequency was only reduced from $p < .002$ to $p < .005$. While the effect of previous conditional probability may not be entirely due to the presence of lexicalized collocations, it is certainly amplified by them, and is undoubtedly due to a few of the 10 function words, especially *of* and *to*, that are strongly associated with certain preceding words.

Word frequency appears not to have much influence on the word forms of function words. There was no effect for the mid/low-frequency function words, even though they represent a range of frequency differences of over ten to one. Nor was there a frequency effect for the function words as a whole, once the difference between the high-frequency words and the rest was accounted for. Since the frequency range for the 10 highest-frequency function words is much less, about three to one, the frequency effect found for those words is suspect. As suggested above, it is likely that the frequency variable encodes individual differences in the combinatorial behavior and form variation of the words rather than a difference in behavior attributable directly to their frequency of occurrence. As a group, however, the highest-frequency function words are more susceptible to reduction than other function words, independently of their higher frequency, as the split variable shows when all function words are analyzed together.

3.3 All words

We now turn to the direct comparison of content words and function words. Results from the analysis of a sample drawn from both classes are summarized in Table 8. The strongest predictor of duration in this analysis is lexical class, the distinction between content and function words. Function words have relatively shorter pronounced durations than content words, after accounting for the effects of the predictability variables and their interactions. After accounting for lexical class, following conditional probability and word frequency are also significant factors. Previous conditional probability is marginally significant. No effects were found for repetition or its interactions.

Table 8. Factors affecting the duration of all words.

factor	ΔR^2 (percent)	F(1, ~1365)	significance
main factors			
content vs function (lexical class)	.96	49.0	$p < .00005$
frequency	.25	12.9	$p < .0005$
frequency^2	.22	11.1	$p < .001$
following conditional probability	.65	33.3	$p < .00005$
previous conditional probability	.11	5.7	$p < .02$
interactions			
lexical class X previous conditional prob.	.24	12.3	$p = .0005$
lexical class X log rate	.05	2.5	$p = .11$

Because effects of the frequency and predictability variables differed for function and content words when they were considered separately, we expect to find that those variables interact with lexical class when content and function words are considered together. Given the strong correlation of lexical class with frequency, however, lexical-class interactions will compete with frequency interactions. One or the other might be significant additions to the regression model; it is less likely that both would be.

There was a strong interaction of previous conditional probability with lexical class. This is expected, because the differences found in the separate analyses were more strongly connected to the lexical class differences than to frequency differences. That is, previous conditional probability was not a factor for content words at all, and its effects on function words were mainly limited to the high-frequency words.

Both the interaction of following conditional probability with frequency (in the positive direction), as well as the quadratic variable frequency^2, were significant factors individually. As with content words, these two factors competed strongly with each other. Frequency^2 was the stronger effect, and is the only factor reported in Table 8, since the interaction of following conditional probability with frequency made no significant further addition to the fit of the model. Thus the stronger frequency and/or predictability effects at higher levels outweighed the lack of a following conditional probability effect for high-frequency function words.

Finally, the interaction between lexical class and frequency was significant when added to the main-effect factors. This reflects the fact that frequency was a strong factor for content words, but not significant for function words, except for the highest-frequency ones. Note that the direction of this interaction (stronger frequency effects for the lower frequency content words) outweighs the shorter duration of the high-frequency function words and stronger frequency effects at higher frequencies for content words. The lexical-class/frequency interaction competes with the conditional-probability interactions. Consequently, it was no longer a significant addition to the model including the previous conditional probability interaction and frequency^2, and is not reported in Table 8.

This sample composed of all word classes is the one most appropriate for reassessing the control variables. While the factors that they represent are known from other studies to affect word durations, it is worth verifying that they remain significant factors in a full model including the probabilistic variables and lexical category. For the model including the significant variables in Table 8, the regression coefficients of the control variables all have significant one-tailed t values. Log rate, accent, and average word length are significant at the level of $p < .00005$; number of syllables, at $p = .0001$; sex, at $p < .001$; final intonational phrase position and sex X log rate, $p < .002$; and age, $p < .05$.

Table 9 summarizes results for the analysis of a sample of the mid/low-frequency words, which excludes the high-frequency words above log frequency -2.1 (above frequency .008),. Recall that these are almost all function words. Lexical class continues to be the strongest factor, confirming that this difference is not due to the special characteristics of the high-frequency function words. Following conditional probability is also a strong factor, reflecting its importance for both content and function words in this sample. Without the influence of the high-frequency function words, the influence of frequency is much weaker, marginal at best. Previous conditional probability is still marginal, not much changed. There were no significant interactions. This was to be expected for the previous conditional probability interaction. The lack of other significant interactions with frequency or of a frequency^2 factor may be related to frequency's marginal significance when high-frequency words are excluded.

Table 9. Predictability factors affecting duration of mid/low-frequency words. There were no significant interactions.

factor	ΔR^2 (percent)	F(1, ~1355)	significance
main factors			
content vs. function (lexical class)	2.77	111.8	$p < .00005$
frequency	.11	4.5	$p < .05$
following conditional probability	1.10	44.5	$p < .00005$
previous conditional probability	.12	5.0	$p < .05$

Since it is only in the mid-frequency region that content and function words mainly overlap, we also analyzed a sample of just mid-frequency words to ensure that the low-frequency words, almost all content words, did not affect the results. The results did not differ essentially from the mid/low-frequency words.

4 Discussion

These results clarify some of the important questions about the relationship between words' frequency, predictability, and articulatory form raised in the introduction. Content and function word durations clearly differ in their sensitivity to probabilistic variables and repetition.

Moreover, high-frequency function words differ from mid- and low-frequency function words. Effects of the variables are compared for these three classes of words in Table 10. **Word frequency**, once contextual predictability factors are taken into account, is an additional contributing factor for content words; it is not a factor for the mid/low-frequency function words; and it is probably not a factor for the high-frequency function words either. Durational effects of **previous conditional probability** are largely limited to high-frequency function words. **Following conditional probability** has the widest application of the probabilistic variables. It is an important factor for content words and mid/low-frequency function words. Although no effect was found in this data, it probably also influences the durations of some high-frequency function words, given the results of Bell et al. (2003). **Repeated** content words, but not function words, have shorter pronunciations than first occurrences, after accounting for intonational and predictability factors.

Table 10. Summary of predictability effects.

	high frequency function words	mid/low frequency function words	content words
word frequency	NS	NS	highly significant
previous conditional	highly significant	marginally significant	NS
following conditional	NS	highly significant	highly significant
repetition	NS	NS	significant

Furthermore, there are differences in the pronounced durations of the three classes of words: High-frequency function words are relatively shorter than mid/low-frequency function words, and the mid/low-frequency function words are relatively shorter than content words, in addition to any frequency or predictability effects. Hence both lexical class and word frequency are required to fully account for variation in word durations.

Effects of frequency on duration are stronger for words which are more frequent. (There is some suggestion that there may also be a similar reinforcing effect between frequency and following conditional probability for content words.) The effect is the reverse of the relationship found by Griffin and Bock (1998), namely that the frequency effect of lexical access was diminished for **highly** predictable contexts. Note also that in their lexical decision study, Segalowitz and Lane (2000) found stronger associations of latency and frequency for **lower** frequency words.

Not all predictability variables affected duration in this data; some made no significant contribution to the fit of the regression models in addition to the variables discussed above. There were no predictability effects on duration from words more distant than the previous or following word. There was no overall effect of conditional probability given both previous and following words, although this does affect the durations of at least the conjunction *and*, as pointed out below in section 4.2. And there was no additional effect of the joint probabilities with the previous and with the following words. This does not necessarily mean that these measures of predictability have no association with word durations. It could mean that the cases where their effect is much greater than the average effect of the accompanying frequency and previous/following conditional probability factors do not occur very often, at least in this data. Nor does it exclude the possibility that there are additional effects of global context that are not captured by these variables, although at present there are no practical ways to quantify such predictability variables for conversational texts.

The finding that when content words are repeated, they are shorter, whereas repeated function words do not shorten, extends earlier research in several ways. They confirm that the repetition effect holds in spontaneous conversational speech. Further, they indicate that the effect is both in addition to the shortening of second mentions by a reduced expectation of accentuation, confirming the results of Bard and Aylett (1999), as well as in addition to shortening by higher contextual predictabilities. Our results indicate that function words are not generally affected by repetition, and that shorter second mentions are at least mainly associated with content words. Verbs, adverbs, and adjectives, as well, consistent with the results of Pluymaekers et al. (2005a). Since Fowler (1988) failed to find repetition effects for form repetition alone (homonyms, lists, etc.) without any informational status difference, this would indicate that either verbs, adverbs, and modifiers are often tied to a system of informational status in some way, or else that there is instead (or, more likely, in addition), an effect of lemma repetition on lexical access.

Does the effect apply to mentions subsequent to the second? At least not substantially, given our results. But the shortening effect is small, so that further shortening by repeated repetitions might well not be detected. This question is related to the repeated-lemma/given-status issue. If given status is the source of repetition reduction, further use of the word would not change its status, so subsequent mentions should not receive additional shortening. Cumulative effects from multiple repetition are more consistent with a repeated lemma source. Further research is needed to clarify these issues.

These results conform to the general principle that word forms have reduced pronunciations when they are more frequent or predictable, the Probabilistic Reduction Hypothesis. But how is this principled variation achieved in speech production? We address this question in the following sections, to the extent that our results are relevant. The relationship between lexical access and articulatory form is taken up in section 4.1. This connection is motivated by the pattern of word frequency effects on duration. A mechanism linking access and form is proposed, based on temporal coordination between the articulatory stream and the planning of utterances at higher levels. Section 4.2 considers the possible production sources of the effects of previous and following predictability on duration, concluding that there are sources of these effects at all levels of speech production planning and execution. Section 4.3 describes the consequences for certain claims about production models, especially the issues about the differential processing of content and function words. We argue here that our results support models in which function words have a special mode or modes of access and in which the access of content words is sensitive to their activation levels, such as the model proposed by Lapointe and Dell (1989). Interestingly, it is the greater sensitivity of content words to frequency that most strongly supports the differential processing of content and function words, and not the finding that function words have shorter relative articulatory realizations than do content words.

4.1 Lexical access, word frequency, and articulation

Lexical access is a likely source of the word frequency effects on articulation summarized above. There are two compelling reasons for this. One is that more frequent words are known to be accessed faster; this is generally attributed to their higher activation (e.g., Griffin & Bock, 1998, and references therein). Another reason follows from our finding that it is content words that are more reduced if they are more frequent, not function words. This suggests a source at the level of lexical access, where the distinction between content and function words is readily available. But for lexical access to be a source for the word frequency effects on articulation, as we mentioned in the introduction, there must be a production mechanism that associates a word's lexical activation with its strength and rate of articulation.

We propose that this association is implemented by a short-term coordination that moderates the pace of articulation when the progress of phonological encoding is slowed. A general motivation for a mechanism like this comes from the need for the production system to maintain temporal coordination between the conceptual/lexical and articulatory temporal streams of speech. More specific support comes from four patterns of probabilistic reduction found here: a word frequency effect in addition to predictability effects, word frequency affecting content words but not function words, the stronger effect of frequency at higher word frequencies, and the lack of an effect of repetition for function words.

The rate at which constructions are assembled and their associated words are accessed and incorporated into a phonologically encoded string is limited by factors such as syntactic complexity and the lexical activation levels of the words and their competitors. The rate at which the articulatory plan is created and executed, on the other hand, is mainly limited by the number and complexity of syllables. It is thus necessary that there be means of maintaining synchrony between planning and articulation. Two mechanisms that help coordinate the two levels are well-known. One is speakers' control of explicitness and rate in different styles of speech, including listener influences. The other is the appearance of disfluencies (pauses, filled pauses, and repetitions, which are also accompanied by longer and fuller articulations) when coordination breaks down.

Our proposal is that there is also a mechanism of fluent speech that helps coordinate lexical access and/or phonological encoding and the execution of the articulatory plan. This would come into play when the specification of the current prosodic unit, most likely the phonological word (Levelt, 1992; Wheeldon & Lahiri, 1997), is slowed, but not so severely to require disfluent adaptations. In order to maintain coordination of the flow on the two levels, the subsequent phasing and/or strength of the gestural articulation of this unit would then be modified slightly, so that it is executed with a longer duration. (Since the phonological unit is provided to the articulator, it is this unit that is affected, not the previous one, which is presumably already underway. In models of availability-based production (Ferreira & Dell, 2000; Frank & Jaeger, to appear), on the other hand, a unit is so slowly constructed that it is not available to the articulation in time, and it is the **previous** stretch of speech where slowing, pausing, etc. takes place to wait for it). There is thus a continuum from fully fluent and rapid speech, through fluent speech that is locally slowed, to the lengthened pronunciations that are part of the disfluency adaptations to incomplete plans that feed into phonetic and articulatory coding. If the base articulation strength is set by long-term accommodations to the conversational context, then the short-term coordination may be thought of as a lengthening mechanism, which shows up in our data as longer pronunciations for less frequent words. This need not imply a direct link between a word's frequency, its access time, and adjustments to phonetic form or articulation rate. Word frequency is not the only factor affecting lexical access time, lexical access time is not the only factor affecting phonological and prosodic encoding, and slower access may not always lead to longer planning times. To accord with the observed lengthened durations for lower-frequency words, it is only necessary that slowed speech plans be more frequent for lower-frequency words than for higher-frequency ones.

Lexical access as the locus of word-frequency effects is supported by the lack of a word-frequency effect for function words: at this level, conceptual and functional differences between content and function words can govern different modes of access. The lack of a frequency effect for function words implies that function words are not usually subject to slow enough access to require accommodation by the articulatory plan, or else that their access does not vary much with respect to frequency. This is not likely to be simply a consequence of the overall higher frequency of function words, since for content words the frequency effect appears to be stronger, at least not weaker, for higher-frequency words. Moreover, if frequency were to affect lower levels of segmental and prosodic organization directly, there is no simple way to exempt forms derived from function words. Thus the lack of a frequency effect supports a different mode of usual access for most function words, an issue we return to below in section 4.3.

Further support for a coordination link between access and phonological encoding and the pace of articulation comes from the finding that reduction effects are diminished for less frequent and /or less predictable content words. The diminution of the frequency effect that Griffin and Bock (1998) found (for **more** predictable words) arises from smaller contextual facilitation for higher frequency words, which are already highly activated, than for less frequent ones. The effect found here cannot come from some sort of similar limitation on the amount of reduction from frequency and/or predictability, either at the lexical level or lower, since it is the less frequent words which undergo less reduction. Nor can the scarcity of collocations among less frequent and less predictable words be more than a minor source of less reduction for these words. Rather, it looks like a ceiling effect in the opposite direction, on lengthening and strengthening, rather than on reduction. In fluent speech, the local slowing of articulation to accommodate higher-level delays would be capped at the lower frequencies and lower predictabilities associated with problems that trigger a disfluency. This limitation on fluent coordination would be consistent with the weakening of the frequency effect for low-frequency content words. There may be other sources for the weakening of the effect, of course. One possibility yet to be explored is that various reduction processes may differ in their sensitivity to frequency and predictability, and that the contexts conducive to the more sensitive ones occur more often in frequent words. On the other hand, reduced frequency effects at lower frequencies are not compatible with a listener-oriented explanation such as selecting certain low-frequency words perhaps unfamiliar to the listener for more explicit pronunciation.

The characteristics of repeated-word reduction also accord with the lexical access and coordination mechanism, whether the source is from repeated lemma access or from given status. Once a word is accessed, either in production or comprehension, its activation level is increased at the conceptual or lemma level. This would facilitate the retrieval of the word's form in subsequent uses. The link with reduced forms would come from the coordination of the progress in planning and the execution of the articulatory plan. The lack of reduction in repeated function words would follow from the restriction to activation-related access to content words. The alternative to priming by lemma access is the proposal of Bard et al. (2000) that the source of repetition reduction comes from given status increasing the lemma activation of second and subsequent mentions. The association with reduced forms would again come from the coordination mechanism. A given-status source of course also accounts for the content-function word difference, by stipulating that function words do not participate in the pragmatic system of new-given status. Note that the restriction of repetition reduction to contexts requiring phrasal planning follows from either lemma priming or given status, if their effects are achieved via a planning-articulation link.

Since the proposed coordination mechanism would normally only come into play for spontaneous connected speech, it would help explain the paucity of duration effects in studies of lexical access in tasks like picture naming. The conclusion that lexical access is the locus of word-frequency and repetition effects on production suggests an important link to comprehension. Models of frequency and predictability effects on lexical comprehension dating back to Morton (1969) or Forster and Chambers (1973) have placed the locus of these effects in the lexical access process. Our results thus provide evidence for shared or at least similar processes of lexical access in comprehension and production.

4.2 Multiple sources for contextual predictability reduction

In this section we consider the sources of the shorter durations associated with greater contextual predictability. We find that there are potential sources of contextual predictability effects at all stages of the planning and execution of utterances. For some of them, there is empirical evidence from previous research; support for others is more indirect and speculative. We argue that no one of the sources alone can account for the observed pattern of effects, and conclude that durational differences attributable to contextual predictability have multiple sources in speech production. The different sensitivities of word classes to previous and to following conditional probability are a further issue. Extrapolating from some particular cases, we suggest that reasons for directional differences in predictability sensitivity may lie in differences in the form and structural combination of words and word classes.

Predictability effects may originate before the level of lexical access. Gahl and Garnsey (2004) and Gahl et al. (2006) showed that predictability from syntactic constructions affected word durations in read sentences. In our analyses, predictability from neighboring words is not distinguished from predictability from syntactic constructions. So some unknown, but probably small portion of contextual reduction found here may be associated with the interaction of syntactic planning and the subsequent activation and retrieval of words from the lexicon. In addition to syntactic sources of predictability, if contextually more predictable lemmas are more strongly activated by preceding or following words, similarly to the effects found by Griffin and Bock (1998), then their forms will be accessed more quickly. Raised activation levels from either syntactic predictability or lemma association could then influence articulations via the coordination mechanism described above in section 4.1. Such sources are not likely to be major contributions to contextual predictability effects, for two reasons. First, they are second-order effects, modifying lexical frequency effects, whereas following contextual predictability affects durations in our results as strongly as or more strongly than frequency. Second, these sources should only apply to content words, given the assumption that their articulation, but not that of function words, is sensitive to frequency. If their contribution were large, we should have found some indication that content words are subject to greater predictability effects.

Other sources of contextual predictability reduction can be loosely characterized as collocational, in the sense of words being more or less closely associated. This can occur at several levels: at the lexical level, at the level of the phonological encoding of prosody, and at the articulatory level.

At the lexical level, lexicalized compounds and other lexicalized collocations increased the contextual effects found in our analyses, to the extent that their components were entered as separate orthographic words by the Switchboard coding. For items whose lexical status is in little doubt, like *ice cream* or *blue jeans*, this is just an artifact of this particular data and its analysis. But there are likely many items whose lexical status varies from person to person, or which may even be in transition, accessible either as a lexical unit or via its component words. It is plausible for this class of collocations that those with higher contextual predictabilities will also be more likely to be accessed as units. Assuming that unit access is usually quicker than access of components, they will be less often subject to slowing by coordination accommodation. The investigation of collocations in section 3.1 indicated that lexical collocations are not a major contribution to following conditional probability effects for content words. This is what one would expect if such collocations are largely limited to higher probabilities. For function words, especially the high-frequency function words, lexical collocations may affect durations more strongly. This issue is addressed below in the discussion on the different sensitivities of word classes to contextual predictability effects.

At the level of phonological encoding, prosody is another potential source of predictability reduction. Predictable collocations, even if not accessed as a lexical unit, will be more likely to be encoded as single prosodic words rather than separate ones, which should usually result in shorter pronunciations. And if predictable sequences are more likely to occur within prosodic phrases and intonational phrases rather than across them, or in longer phrases, then their articulations will in some instances be more reduced than less predictable sequences. This would follow from their not being subject to initial or final lengthening or by the shortening of units within larger domains (Fougeron & Keating, 1997). This hypothesis of course needs empirical validation, which is beyond the scope of this study. We did examine whether some necessary preconditions for this mechanism exist at the level of the intonational phrase, since words in the dataset were coded for length of the intonational phrase (IP) containing them. Indeed, IP length and following conditional probability are significantly correlated with duration in the appropriate directions. Further, for the content-word sample excluding high-frequency homonyms (and also excluding *kind of* and *lot of*), IP length is a significant additional factor ($F(1, 1205) = 6.3$, $p = .01$), and reduces the contribution of following conditional probability.

At the articulatory level, certain frequent word sequences may have stored articulatory plans, automatized by repetition, as suggested by Bybee and Hopper (2001) and Bybee and McClelland (2005). Presumably the contribution of this mechanism to contextual predictability reduction would be limited to higher-frequency words. Moreover, it cannot be a dominant source, since if very frequent repetition were the source of such plans, joint contextual probability would have been a stronger factor.

In addition to these collocational and prosodic mechanisms, a direct influence of predictability on lower levels cannot be ruled out, although it suffers the disadvantage of a lack of independent motivation. Biasing of the selection of phonetic exemplars, as proposed by Pierrehumbert (2002), is perhaps one mechanism that could link contextual predictability directly to articulatory realization. Direct-influence models are discussed further below in section 4.3.

So far we have not addressed the different sensitivities of content and function words to predictability from neighboring words, especially predictability from the previous word. Why should function words with higher previous conditional probability be shorter, but not content words? First of all, differences in contextual predictability are not just content-function differences. It is not the case that all function words have shorter pronunciations when they are more predictable from the previous context; it is just the most frequent function words. This is only an indication of differences that likely extends to much smaller classes of words. Among the 10 most frequent function words, Bell et al. (2003) found that some are sensitive both to previous conditional probability and to following conditional probability (*the, to, it*). Some are sensitive only to following conditional probability (*a, in, that, you*). And *I* and *and* are sensitive only to previous conditional probability. Function words of course have different dependencies with surrounding structures and this can affect how their forms are accessed (Alario & Caramazza, 2002; Eberhard, Cutting, & Bock, 2005). Whether this might be related to the differences found here is unknown. Different effects of predictability probably also exist for less frequent function words and content words. (See, for example, the differences found by

Pluymaekers et al. (2005a) for seven Dutch adjective/adverbs). If so, one must conclude that differences in the direction of predictability sensitivity reflect the fact that words or classes of words typically occur in different kinds of constructions. A particularly striking instance of a construction-specific effect is given by *and*, one of the few words to be sensitive to the bilateral conditional probability given both previous and following words. It occurs frequently as a marker of binomial constructions such as *now and then*, *mom and dad*. *And* is pronounced shorter in binomial constructions than elsewhere, and in binomials, it is shorter when more predictable from both the surrounding words (Bell et al., 2003). Even though no overall significant contribution from bilateral conditional probability was found in the present data, certain other words are likely to be affected if they occur in similar constructions; an obvious candidate is the disjunction *or*.

From this perspective, the patterns of sensitivity to measures of contextual predictability displayed in Table 10 stem from preponderances of certain kinds of forms among very frequent function words, other function words, and content words, and from the structures they typically occur in. Elucidating the reasons for the differences will require more detailed word and class analysis, combined with more detailed and specific measures of reduction than the global one of duration employed here. Some factors that could contribute to the different patterns come to mind. For frequent words and combinations, the influence of possible lexicalized collocations with previous or following words is an obvious candidate. Form differences are another. As pointed out earlier, very frequent function words are monosyllabic and have frequent reduced alternate forms. They often include alternate forms lacking onsets for the words with obstruent onsets (*the, that, to*). This makes them vulnerable to reduction when closely associated with previous words, and some of them at least do occur in constructions with such associations. Resulting reduced forms suggest an earlier stage along the path that has produced lexicalized cliticizations of *is*, *have*, *not*, etc. Content words, on the other hand, generally contain at least one full or stressed syllable, which in English is most frequently the first syllable (Cutler & Carter, 1987). The lack of an effect of previous conditional probability for content words might reflect a relatively lower incidence of close association with preceding words. Even if it does not, the ends of content words are more vulnerable to reduction than their onsets, so that shortening is more likely in cases of close association with following words. Other explanations for the predominance of following predictability involving the time course of planning, such as the tentative “involvement-in-planning” suggestion of Pluymaekers et al. (2005a), are yet to be considered.

4.3 Probabilistic reduction and cognitive models of speech production

Our results add further support to the different modes of lexical access of content and function words proposed by Garrett (1975) and Lapointe and Dell (1989). (See also the proposal of Gordon and Dell (2003) that lexical activation of heavy verbs depends mainly on semantic cues whereas activation of light verbs (e.g. *go*, *do*) depends more on syntactic cues, which could be extended to content-function differences.) Models like that of Stemberger (1985), in which content and function words alike are accessed by network activation, are not compatible with the strong effect of frequency on the duration of content words, and the lack of one for function words. Nor are they compatible with the restriction of repetition reduction to content words. There are undoubted differences between content and function words in form and function, of course, which Stemberger invokes to explain the speech error and aphasia motivations for their differential access. These differences, however, some of which are noted above in section 4.2 and in the following paragraph, are generally of the sort that would lead to greater reduction of higher frequency function words, not less. They also have no apparent connection with repetition. It is thus difficult to see how they could account for different sensitivities to frequency and repetition of content and function words. Note, however, that nothing about the effects of probabilistic reduction addresses the different ways that function words are accessed in the models of Garrett (prespecification in syntactic templates) and of Lapointe and Dell (access by a feature-lookup procedure).

The shorter articulations of function words compared to content words of equivalent frequency and predictability also appears to support a different mode of access of content and function words. This difference is compatible with differential access, but it does not provide strong support for it, mainly because here the representational, form, and structural differences of function words do provide a plausible alternative sources of the duration differences. (Differential access also would not account for the articulatory duration differences between the high-frequency function words and the rest.) One such source is that function words have common alternative reduced forms. Alternative lexical forms are usually assumed for a few words like *the* (ði~ðə) and *a* (eɪ~æn), but they probably exist for many others. For example, *a* and *of* are mostly pronounced ə, ə/əv, and in faster speech almost always so. Content words containing the lexical segments eɪ and ɑ such as *day*, *say*, *lot*, *rob*, etc. have no such alternants. This is particularly particularly characteristic of the high-frequency function words, but occurs with others as well, e.g. *because*, *unless*, *if*, *between*. This would allow contextual choices among forms of varying degrees of fullness and/or reduction at the level of lexical selection. In analyses such as this one, in which control of form is based on assumed single lexical representations, to the extent that function words overall have more reduced variants, they will end up with shorter relative pronunciations than content words. Note that this need not presume that no content words have such variants; some do, e.g. *regular*, *probably*, *hundred*. Nor does it presume that content words in general cannot sometimes be highly reduced, as for example in the pronunciation of *think* in our corpus cited earlier. (See Johnson (2004) for an extended discussion of reduction of content words.) Another contributing factor may be the high proportion of function words which have monosyllabic, unstressed, and often reduced forms. ("Reduced" is used here in the specific sense of not being a full syllable, like the second syllable of *pin**na**cle* versus that of *pi**no**chle*). Length variation, especially that involving vowel reduction and the deletion of onsets, vowels, and codas is found more often in unstressed syllables. This may encompass the entirety of many function words, but only a part if any of many content words, so that any shortening effects are relatively greater for function words.

What are the consequences of the present results for cognitive production models that account for probabilistic reduction by a direct influence of frequency and predictability on articulation? To the extent that such proposals only assert that there must be some connection between words' frequency and predictability and their articulatory realization, they obviously accord with our conclusions. If, however, the claim is that some representation of the frequency and predictability of words is passed through (or in parallel to) the levels of lexical access and sequencing and of phonological encoding to the phonetic and articulatory levels, then there are some potential contradictions. The application of frequency and of predictability together at the same level (e.g. the "simultaneous" application of Pluymaekers et al., 2005a) is at odds with frequency (and repetition) affecting content words but following conditional probability affecting function words as well as content words. Even if separate parameters for frequency and predictability are assumed, a motivated mechanism is lacking that could distinguish content and function words at a level which most models presume to be composed of entities close to articulatory forms. Another possible mechanism for a direct-influence model is that word frequencies and predictabilities are in fact reflections of the frequencies and sequential predictabilities of articulatory units, perhaps syllables, syllable-like units, or in some cases short syllable sequences. Then perhaps the higher activation and more rapid access of the more frequent and more predictable units would lead to weaker and shorter articulations. If this were the case, however, one would expect that there would be ceiling effects that would reduce probabilistic effects at higher frequencies and predictabilities. This is the opposite of the effect that was actually found.

Another class of alternative explanations is provided by the accommodation of speakers to the needs of their audience, whereby more redundant elements, being more retrievable from context, are allowed to be less fully articulated (e.g. Lindblom, 1990). Speakers certainly adjust their speech to suit different audiences and situations in multiple ways, including choices of explicit or elliptical forms, slower or faster rates of speech, and reduced or full pronunciations. Bard et al. (2000) point out that these adjustments are computationally complex, given the demands of speech production planning, so they take place slowly and imperfectly. Hence they may not be not good candidates for the source of the local and rapid form variations characteristic of probabilistic reduction. In any case, more details are needed about the nature of audience-oriented mechanisms and their place in the production process in order to assess whether they can account for the patterns of probabilistic reduction that were found here. In particular, it would appear intuitively that predictability from the previous word would be a much more prominent factor if the audience's needs were paramount.

Finally, it is possible to recast probabilistic reduction in terms of redundancy or informativeness, and to presume that the production process is governed by smoothing mechanisms that ensure the even distribution of informativeness over the time course of speech (Aylett, 1999; Levy & Jaeger, 2007; Van Son, Koopmans-van Beinum, & Pols, 1998). These may be just different ways of characterizing the same phenomenon. If they are truly different, then the mechanisms proposed here provide a basis for comparison with possibly distinct redundancy-smoothing mechanisms.

We conclude that four general hypotheses about the sources of probabilistic reduction in speech production follow from the patterns of association of form duration with measures of frequency and predictability:

- Longer durations for less-frequent content words have their main source in their lower lexical activation levels, which lead to slower retrieval from the lexicon.
- Function words have a privileged means of access.
- Shorter durations for more contextually predictable words have multiple sources, including those related to lexical activation and to the grouping of words at the lexical and prosodic levels.
- The reduction of second mentions stems from an augmentation of the lexical activation level of a word.

References

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: evidence from French. *Cognition* **82**, 179-223.

Aylett, M. P. (1999). Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and syllabic duration. *Proceedings of the 1999 International Conference on Spoken Language Processing* (pp. 289-292).

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, **47**, 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllable duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, **119**, 3048-3058.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language,* **28**, 14-36.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language,* **42**, 1-22.

Bard, E. G. & Aylett, M. P. (1999). The disassociation of deaccenting, givenness, and syntactic role in spontaneous speech. *Proceedings of the 1999 International Conference on Spoken Language Processing* (pp. 1753-1756).

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, **113**, 1001–1024.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, **31**, 163-186.

Bock, K. & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In Gernsbacher, M. A. (Ed.), *Handbook of psycholinguistics* (pp. 945-984). New York: Academic Press.

Brill, F. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics* **21**, 543-566.

Browman, C. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. & Beckman, M. E. (Eds*.), Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge: Cambridge University Press.

Bush, N. (2001). Frequency effects and word-boundary palatalization in English. In Bybee, J. & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 255-280). Amsterdam: Benjamins.

Bybee, J. L. (2000). The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. & Kemmer, S. (Eds.), *Usage-based models of language* (pp. 65-85). Stanford, CA: CSLI.

Bybee, J. & Hopper, P. (2001). Introduction to frequency and emergence of linguistic structure. In Bybee, J. & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 1-24). Amsterdam: Benjamins.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* **22**, 381–410.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, **37**, 575–596.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, **14**, 177-208.

Caramazza, A. & Berndt, R. S. (1985). A multicomponent deficit view of agrammatic Broca's aphasia. In Kean, M.-L., ed., *Agrammatism*, pp. 27-63. Orlando:Academic Press.

Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary.

Computer Speech and Language, **2**, 133-142.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **29**, 416-431.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, **5**, 313-349.

Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, **42**, 287-314.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, **112**, 531-559.

Ferreira, V. S. & Dell, G S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* **40**, 296–340.

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

Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, **12**, 627–635.

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

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, **101**, 3728–3740.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, **31**, 307–319.

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

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

Frank, A. F. and Jaeger, T. F. (to appear). Speaking rationally: Uniform information density as an optimal strategy for language production. Proceedings of the 30th Meeting of the Cognitive Science Society.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, **80**, 748–775.

Gahl, S., Garnsey, S. M., Fisher, C., & Matzen, L. (2006). "That sounds unlikely": Syntactic probabilities affect pronunciation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1334-1339).

Garrett, M. F. (1975) The analysis of sentence production. In Bower, G. H. (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 9 (pp. 133-177). London: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In Bower, G. H. (Ed.), *Language production*, Vol. 1 (pp. 177-220).

Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the IEEE ICASSP-92* (pp. 517-520). San Francisco: IEEE.

Goodglass, H. & L. Menn. (1985). Is agrammatism a unitary phenomenon? In Kean, M.-L., ed., *Agrammatism* (pp. 1-26). Orlando: Academic Press.

Gordon, B. & Caramazza, A. (1982). Lexical decisions for open and closed class words: Failure to replicate differential frequency sensitivity. *Brain and Language* **15**, 143-160.

Gordon, J. K. & Dell, G. S. (2003). Learning to divide the labor: an account of deficits in light and heavy verb production. *Cognitive Science* **27**, 1-40.

Greenberg, S. (1997). Switchboard transcription system. Labelers' manual, revision of 19 February 1997. ms.

Greenberg, S., Ellis, D., & Hollenback, J. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. *Proceedings of the 1996 International Conference on Spoken Language Processing* (pp. 24 –27).

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

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

Hawkins, S. & Warren, P. (1994). Implications for lexical access of phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics,* **22**, 493-511.

Herron, D. T. & Bates, E. A. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *Journal of Memory and Language,* 37, 217-239.

Hocking, R. R. (2003). *Methods and applications of linear models: Regression and the analysis of* *variance, 2nd Ed*. New York: Wiley and Sons.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (Ed.), *Current progress in historical linguistics* (pp. 96–105). Amsterdam: North Holland.

Jespersen, O. (1924/1965). *The philosophy of grammar*. New York: W. W. Norton.

Johnson, K. (2004). Massive reduction in conversational American English. In Yoneyama, K. & Maekawa, K. (eds.), *Proceedings of the 1st Session of the 10th International Symposium on Spontaneous Speech: Data and Analysis.* Tokyo: The National International Institute for Japanese Language (pp. 29-54).

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J., & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.

Kello, C. T. & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **26***,* 719-750.

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

Lapointe, S.G. & Dell, G. S. (1989). A synthesis of some recent work in sentence production. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 107-156). Dordrecht: Kluwer.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Leslau, W. (1969). Frequency as determinant of linguistic change in the Ethiopian languages. *Word*, **25**, 180–189.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes, and representations. *Cognition*, **42**, 1-22.

Levelt, W. J. M. (2002). Phonological encoding in speech production: Comments on Jurafsky *et al.*, Schiller *et al.*, and van Heuven & Haan. In Gussenhoven, C. & Warner, N. (Eds.), *Laboratory phonology 7* (pp. 87-99). Berlin/New York: Mouton de Gruyter.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, **22**, 1-75.

Levy, R. & Jaeger, T.F. (2007). Speakers optimize information density through syntactic reduction. In

B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (pp. 849-856). Cambridge, MA: MIT Press

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

Lindblom, B. (1990**)**. Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle. W. & Marchal, A. (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer.

Liu, H., Bates, E., Powell, T., & Wulfeck, B. (1997). Single-word shadowing and the study of lexical access. *Applied Psycholinguistics,* **18**, 157-180.

Manning, C. & Schütze, H. (1999). Collocations. Chapter 5, *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

MacWhinney, B. (2001). Emergentist approaches to language. In Bybee, J. & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 449-470). Amsterdam: Benjamins.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76**, 165-178.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In Cole, J. S. and Hualde, J. I. (Eds.), *Laboratory phonology 9* (pp. 201-228). Berlin/New York: Mouton De Gruyter.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, **17**, 273–281.

O'Shaughnessy, D. (1992). Automatic recognition of hesitations in spontaneous speech. In *ICASSP-92*, Vol. I (pp. 593–596). New York: IEEE.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In Gussenhoven, C. & Warner, N. (Eds.), *Laboratory phonology 7* (pp. 101-139). Berlin/New York: Mouton de Gruyter.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, **62**, 146–159.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, **118**, 2561-2569.

Rhodes, R. A. (1992). Flapping in American English. *Proceedings of the 7th International Phonology Meeting* (pp. 217-232). Turin: Rosenberg and Sellier.

Rhodes, R. A. (1996). English reduced vowels and the nature of natural processes. In Hurch, B. & Rhodes, R. A. (Eds.), *Natural phonology: The state of the art* (pp. 239-259). The Hague: Mouton de Gruyter.

Schuchardt, H. (1885). *Über die Lautgesetze: Gegen die Jungrammatiker*. Berlin: Robert Oppenheim. Excerpted with English translation in Vennemann, T. & Wilbur, T. H. (Eds.). (1972). *Schuchardt, the Neogrammarians, and the transformational theory of phonological change* (pp.1-72). Frankfurt: Athenaum Verlag.

Segalowitz, S. J., & Lane, K. C. (2000). Lexical access of function versus content words. *Brain and Language*, **75**, 376–389.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, **42**, 213-259.

Shattuck-Hufnagel, S., & Ostendorf, M. (1999). Posh labeling guide—version 1.0. ms.

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

Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 143-186). London: Erlbaum.

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America*, **107**, 1697-1714.

Van Son, R. J. J. H., Koopmans-van Beinum, F. J., & Pols, L. C. W. (1998). Efficiency as an organizing principle of natural speech. *Proceedings of the 1998 International Conference on Spoken Language Processing* (pp. 2395-2398).

Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, **37**, 356-381.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, **15**, 1-95.