The role of context in lexical recognition

Jacqueline Palmer

Introduction

The nature of the mental lexicon and its stored forms remains an open and much disputed question. While the traditional view has suggested single abstract phonological forms for each lexeme, there has been much recent evidence to the contrary that requires a model of the lexicon that allows much more information to be encoded on several lexical representations per lexeme. There have been many recent studies on a variety of phonetic aspects that seem to indicate that lexical forms need to be encoded in a more detailed manner to be able to account for listeners' great sensitivity to the individual variation that is displayed in everyday speech (e.g., Bybee, 2001; Palmeri, Goldinger, & Pisoni, 1993). Subjects in many experiments have been able to attend to fine-tuned phonetic detail that distinguishes one talker from the next but traditionally has not been included in the abstract lexical representation, such as individual variation in voice-onset-time (Allen & Miller, 2004). These recent studies have suggested that these phonetic details be encoded in the representation of the stored lexical forms as listeners' recognition of these details aided in and sped up lexical recognition.

One question that remains is the status of utterances of words that are phonetically reduced; i.e., that are missing segments that would otherwise be included in a more careful pronunciation of the same word. The studies mentioned above used pronunciations of words that indeed differed from speaker to speaker in a given property, but in both cases the stimuli consisted of words that contained all of the segments that would exist in a careful pronunciation. Thus, it remains unclear whether phonetically reduced forms may be simply another part of interspeaker variation that requires representation in the lexicon just as forms varying in the dialect or gender of the speaker do. This leads one to wonder which phonetically detailed productions listeners remember, if listeners and speakers do indeed retain phonetically detailed representations in the lexicon. While simple accounts of lexical access would likely have heavily reduced conversational tokens be activated by a single abstract representation during recognition, it could be that these reduced forms are retained in memory to be deployed during the recognition of reduced versions of words.

This study seeks to shed light on the nature of the perception of phonetically reduced forms in conversational speech; specifically, the factors that lead to listeners' difficulty in perceiving reduced forms. This question will be examined in the below sections as follows. The next section will review the previous literature on this topic and will discuss relevant studies that have challenged the traditional view of lexical representations, and how models of lexical representations have adjusted to accomodate these results. I will then discuss the main experiment, Experiment 1, which was a word-monitoring task examining the role of surrounding context in deciphering phonetically reduced forms. The creation of the materials and the design of the experiment are discussed at length in this section, as well as the preliminary results of the data collected from this experiment. Following this I discuss a follow-up study to the first experiment, which was a cloze task examining subjects' ability to predict the target words of the experimental set of stimuli given the words and phrases present in the surrounding context sentence into which it was spliced. The materials and the experimental design are discussed in this section, followed by a discussion of the results of the first experiment given the results of Experiment 2. A general discussion of the significance of these findings and proposals for changes to current models of lexical recognition will be given at the end of this paper. The text of the stimuli presented to subjects in Experiments 1 and 2 are given in the Appendix.

Previous studies

The claim that there are multiple representations of each lexeme in the mental lexicon is not a new one; it continually gains support from present studies that I will now briefly discuss. Listeners have been shown to be so sensitive to slight differences between the pronunciation of forms by different talkers that they are able to learn the difference between two talkers' voices and generalize these talker-specific patterns to novel utterances. One example of this effect was shown in Allen and Miller (2004); this study examined individual differences in voice-onset-time (VOT) and showed that listeners were able to learn the VOT patterns specific to a given talker, and were able to demonstrate their learning on novel words from the same talkers on which they were not trained. It is thus well established that multiple representations may be necessary to account for effects such as these. However, most studies of this sort are done on "indexical properties," a term I am borrowing from McLennan and Luce (2005). They describe "indexical properties" as phonetic properties of a form that enable the listener to distinguish between speakers, as opposed to phonemic properties that distinguish lexical items from each other. Previous studies have demonstrated that multiple representations in the mental lexicon may be necessary to explain these effects with forms that differ slightly in an indexical parameter, but none of these accounts cover phonetic reduction as a variable for which multiple representations might be necessary.

This leads one to wonder what the status of the reduced form is in the mental lexicon. Though listeners can store multiple forms of the same lexeme depending on other indexical properties, like speaker rate and average vocal pitch as described in McLennan and Luce (2005), one wonders how listeners can recognize forms that are missing some of the segments necessary to arrive the stored form in the mental lexicon, whatever shape the form takes. It is important to note that phonetically reduced forms (like the ones considered in this study) are basically unintelligible without some surrounding context. This was examined in a pilot study preceding the current study, where I played some reduced forms from the corpus from which the experimental stimuli are taken and asked a group of six listeners to attempt to guess the word that was played to them. The words that were chosen were words that met similar criteria as the target words in the experimental stimuli: the words chosen were multisyllabic and fairly frequently used conversational items but not function words or proper names. All of the forms were phonetically weakened in some way; most were missing one segment from the canonical pronunciation or featured some type of assimilation. For example, "always" took the form [a.lɪs] and "getting" [gɛ.niŋ]. The highest accuracy score for this test was 33%, and most of the other subjects scored much lower.

There have been some accounts that show that preceding context can affect perception and can likely aid the listener in word recognition. Johnson (1995) outlines several studies that examine the role of a phonetic property in a carrier phrase on a target within a carrier phrase. While these studies do exhibit context effects, the properties that the investigators consider are indexical properties (namely, F0 and vowel formants) that do not result in the elision of segments, and the question of reduced forms is still unclear. One other problem presented by the role of context in perception and recognition (of reduced forms and more generally) is that of modeling: most current models of lexical recognition will not permit information about an entire sentence, for example, to feed back down to the level responsible for recognizing words.

The current study was set up to examine whether the style of pronunciation of the preceding context will affect the perception of reduced target forms and will at least partially explain why reduced forms can be perceived with ease in running speech but not in isolation. I examine whether or not the difficulty in understanding reduced forms can be partially explained by a mismatch in pronunciation style, and discuss how this may be considered and represented in models of lexical recognition.

Experiment 1

Even though pilot studies showed that the reduced forms are almost completely unintelligible when in isolation, people have little trouble dealing with phonetic reduction in conversational speech. The very forms that were unintelligible to subjects in isolation were pronunciation variants that one encounters every day in casual speech. Thus it is clear that context aids greatly in the comprehension of reduced forms. Experiment 1 was set up to determine the cause of subjects' difficulty in perceiving phonetically reduced forms. I set up a word monitoring experiment where subjects were instructed to react as quickly as possible to a target word spliced into various context sequences. One possible outcome is that pronunciation variants that are close to the canonical pronunciation will always be easier to recognize than phonetically reduced, casual forms, regardless of the style of speech in the surrounding context. We would then expect subjects to react more slowly to reduced forms than canonical forms in all cases. However, it could also be that the context in which the reduced variant appears is the key to subjects' ability to process them; perhaps subjects' great proficiency at perceiving reduced forms in context stems from the fact that they expect a more reduced variant of a given word when it is uttered in a sequence that features other reduced variants. In this scenario, the hearer's difficulty in understanding reduced forms would be mostly due to encountering a reduced form in a more formal, more fully articulated context. Here, reaction times for reduced forms would be greater than these canoncial forms only when the reduced form occurred in a nonreduced context. A significant interaction between the style of pronunciation of the context sequence and the style of pronunciation of the target word would indicate that it is the mismatch between context and target word that leads to difficulty in comprehension of reduced forms, and not inherent unintelligibility on the part of the reduced forms.

Method

Participants

Thirty-two undergraduate and graduate students at University of California, Berkeley participated in this experiment. They were compensated at the rate of ten dollars per hour of participation. No subject reported any hearing problems.

Materials

I constructed a list of sixteen words of three or more syllables taken from the Buckeye corpus (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). The words that were chosen exhibited a large degree of variability in pronunciation in their appearances in the corpus. I extracted four pronunciations produced by a single talker for each word. The sixteen words came from a total of eleven talkers; subjects heard each individual talker no more than twice in the experimental set. I selected two pronunciations that were pronounced like or very close to a dictionary pronunciation of the word, which I term "full" pronunciations, and two pronunciations that exhibited some amount of phonetic reduction, which I term "reduced". In some cases there were no examples of the dictionary pronunciation; in these cases I attempted to choose two pronunciations that were 1) identical in transcription, if possible, and 2) as close to the dictionary pronunciation as possible. In other cases, there were no two identical reduced pronunciations; here, I attempted to select two pronunciations that were as similar as possible in transcription. The context in which each target word occurred was also extracted from the corpus; these strings varied in length from four to eleven words, where the target word occurred anywhere except the beginning of the string. Finally, a set of four target words and four context sequences for each of the sixteen target lexical items was compiled, and each target word was spliced into both context sequences of the opposite pronunciation category and into the context sequence of its own pronunciation category in which it did not occur.

Table 1: Context-target word pairings for experimental item *similar*, all produced by talker S32. The full form from Full Context 1 is spliced into the second full context and both of the reduced contexts. The target word never occurs in the context sequence from which it came in the corpus.

Pronunciation style	Context sequence
Full Context 1	Because nobody is more *similar* to you than that sibling
Full 2	Is extraordinarily __um, so what she says might
Reduced 3	Is often so __to where I'm coming from
Reduced 4	We were able to communicate in, uh, very __fashions

Table 2: An example of the nature of the full and reduced pronunciations of an experimental item, all produced by talker S32. The full variations are not completely identical, but are clearly distinct from the reduced variations, as the full variations possess one more syllable than the reduced variaitons.

Pronunciation style	Phonetic transcription
Full 1	[sɪ.mɪ.lar]
Full 2	[sa.ma.lar]
Reduced 1	[sɪm.lər]
Reduced 2	[sm̩.lər]

To illustrate, each full target word was spliced into both reduced context sentences and into the full context sentence in which it did not occur. This resulted in a total of twelve target word-context sentence pairings for each word set. Table 1 illustrates this cross-splicing process for the word set *similar*.

Six of the sixteen word sets contained a defective target word and context sentence that did not satisfy all of my selectional criteria, and a proper replacement could not be found in the corpus; this target-context pair was deleted from the word set, and thus the word set contained only three target words and context sentences. Two different randomized lists of fifty-two of the possible pairings were constructed, to which alternating subjects were assigned; each list ensured that no subject heard a given target word in the same context sequence more than once and also created a balance across the two groups in case of the defective word sets. Subjects did monitor for the same target word, in varying pronunciations, multiple times.

In addition to the fifty-two experimental target-context pairings, a distractor set of thirty-two sequences was constructed (see Appendix). These sequences were taken from the same talkers as in the experimental set; there were no distractor items from novel talkers that were not used in the experimental set. In sixteen of the thirty-two sequences, the target word was present; in the other sixteen, the target word that would be presented to the subject did not occur in the sequence. Each of these sets featured a "zero-edit" condition where no splicing was performed to ensure that any noise due to the splicing in the experimental set was not being used as a cue when monitoring for the target word. Each also featured an "edit" condition where a word other than the target word was spliced from a different location in the corpus. Every subject was presented every item from the distractor set. A total of eighty-four sequences were then presented to each group of subjects.

The stimuli were created and spliced using Wavesurfer v. 1.8.5 (Sjolander & Beskow, 2000) and Praat v. 4.4.26 (Boersma, 2001). The stimuli were stored at a sampling rate of 16 KHz.To eliminate telltale noise at the splice boundaries, some small portions of the target and context sound files had to be deleted or set to silence, usually only the length of several glottal pulses. In an attempt to have the spliced sequences and transitions sound as natural as possible, I manipulated the pitch contours of the target words and parts of the context sequences so that the target words and context sequences of each word set had roughly the same pitch level and contour. This was done by manually altering pitch pulses at splice boundaries between the context sentence and the target word so that the contour was smooth, using the PSOLA (Pitch-Synchronous Overlap and Add) method in Praat (Boersma, 2001). In most cases, the level of pitch within each talker's speech was quite stable; with some talkers, however, some context sentences were much lower in pitch than the others in the word set, and this created problems when trying to cross-splice the target words. To resolve this, I made the pitch level of the outlying context sentences and target words equal to the average of the more normal context sentences in each word set.

Procedure

Participants were tested either individually or in groups of up to three. All participants listened to the sequences using headphones; each testing station was determined before each test that its speakers played at a comfortable volume and at the same level as the other testing stations. The target word appeared to subjects in the center of their screen for 5 s before the target word was played in a context sequence. Subjects were instructed to press a button when they heard the target word. At the end of the audio sequence, subjects received a feedback message that simply made them aware that their response was recorded, and then a new trial began immediately afterward. The lack of time between the end of the audio sequence and the feedback message caused some problems for some subjects; in the cases where the target word occurred at the end of the sequence, subjects only had a window of time to respond as long as the target word itself. Some subjects reported that they were able to recognize the target word but were cut off by the program, when it had already moved on to the next trial. Response time was measured from the onset of the target word. Subjects were encouraged in their instructions to react as quickly as possible, but also to be certain that they correctly heard the word that appeared to them on the screen. The entire test was created and run with the E-Prime program (Schneider, Eschman, & Zuccolotto, 2002), a software suite for psychological experimentation that also performed preliminary statistical analysis, to be discussed in the

Table 3: Number of zero responses in each listening condition

Context	Target	Missing	Total	Percent
Full	Full	4	415	1.0
Full	Reduced	17	429	4.0
Reduced	Full	6	400	1.5
Reduced	Reduced	17	416	4.1
Overall	Overall	44	1660	2.7

Table 4: Mean reaction times for each listening condition, in ms

	Full Target	Reduced Target
Full Context	505	531
Reduced Context	543	533

Results section, below.

Results

Reaction times over or below three standard deviations from the mean were replaced by the cutoff value. This procedure affected 1.2% of the data.

All zero responses that resulted from subject error in the experimental set were removed from the reaction times. This affected 44 out of 1660 trials in the experimental set, or 2.7% These were removed from the data set before it was analyzed. The problem I mention above about the lack of response time between the end of the audio sequence and the feedback message may have been the cause of some of the zero reaction times; however, as they only affected a small percentage of the trials I will not consider it a significant problem. Table 3 shows the number of zero responses in each listening condition.

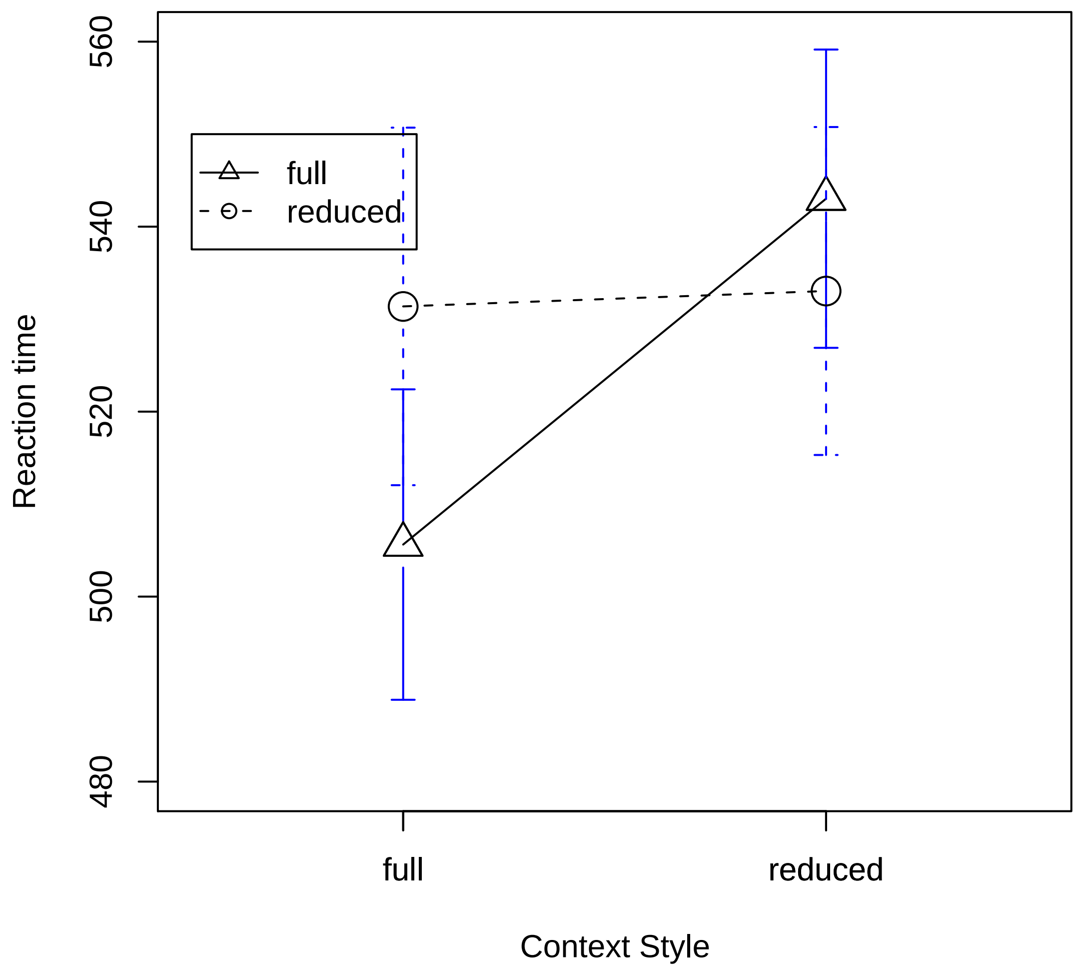
There was no main effect due to the style of pronunciation of the target word [$F_1(1,31) = 1.06$, $p = .31$]. Though this result was not statistically significant, there was a numerical trend for the reduced targets to be perceived more slowly overall, regardless of context style of pronunciation: the mean reaction time for full targets was 524 ms, compared to 532 ms for the reduced targets.

The effect of context style was also not a significant factor [$F_1(1,31) = 2.69$, $p = .11$]. The mean reaction time in full context listening conditions was 518 ms, compared to 538 ms in the reduced context listening conditions.

There was a significant interaction of context style and target style [$F_1(1,31) = 4.34$, $p = .045$]. This significant interaction is shown in Figure 1 and Table 4.

Table 4 shows the mean reaction times in each of the four listening conditions: Full Target - Full Context, Full Target - Reduced Context, Reduced Target - Reduced Context, Reduced Target - Full Context. As can be seen from the table, there is a very small dif-

Figure 1: Interaction of context style and target style. Word monitoring times for words pronounced with a high degree of pronunciation reduction are plotted with circles. Monitoring times for words with the "full" pronunciation style are plotted with triangles.



ference between the reaction times on the reduced targets in the two context conditions. This seems to indicate that the reduced target words are, on the whole, slightly more difficult for the subject to process. Yet the mean reaction times and the significant interaction between context style and target style, as seen in Figure 1, show that the pronunciation style of the preceding context seems to play a role in the subjects' perception and processing of the target words. It is very interesting that not only are the reduced targets processed slower when preceded by a mismatching (full) context, but that full targets are also processed slower when preceded by a reduced context. The reaction time difference in the full context condition is indeed smaller than the reaction time difference in the reduced context condition, but it is still striking that reduced forms are not simply always more difficult to understand and that full forms are sometimes processed more slowly than reduced forms in a mismatching context. It would seem more likely that the full form that carried more distinctive phonetic information to aid the subject in lexical recognition would always be easier for the subject to process. A form with less information could potentially be more easily confused with another form, or at least could potentially be intially more confusable with other lexical items. Marslen-Wilson (1987) defines the class of lexical items that are all equally plausible as a candidate for a lexical item that is being accessed at a given "recognition point" (rather, some location partially through the word) as a "word-initial cohort," as a given lexical item is usually only distinctive from other forms by the end of the word, or after a number of segments. A full target form has greater chance for carrying more distinctive segments or other information that would give it an earlier recognition point, so it would seem that all full forms should be processed more quickly. The fact that a significant interaction was found between the pronunciation style of the context and the target, however, seems to present another explanation for this problem. It is the mismatch between the context and the target that causes the subjects' auditory recognition to slow down, because the phonetic features of the context have likely led the subject to expect a pronunciation of the target word that was similar in style to pronunciations that preceded it. This is what causes difficulty in perceiving not only reduced forms, but also full forms in reduced contexts. This suggests that reduced forms may not be inherently more difficult to perceive than full forms, and that reduced forms may require a status in the mental lexicon equal or nearly equal to full forms if reduced forms can be perceived as quickly as they are in reduced contexts.

Experiment 2

To ensure the accuracy of the results of the first experiment, I administered a cloze task (Taylor, 1953) to examine the predictability of the target words used in Experiment 1 given the specific words and scenarios in the context sequence. The materials used here, which were used in Experiment 1, were taken from conversations in the Buckeye corpus; while using real speech data was the most appropriate choice for this study and while it has many advantages for studies of this type, one important disadvantage was not being able to control the nature of the context sequence. Consequently, some instances of the target word were likely more predictable given certain words that preceded them, which may have obscured the reaction time effect that I examined.

Method

Participants

Four undergraduates at the University of California at Berkeley participated in this experiment. Given the very simple and short task, these subjects were not compensated for their participation.

Procedure

Participants were given a type-written document of all of the context sequences in the experimental set with a blank in the location where the target word would have appeared. The sequences were put in random order and then reordered further to ensure that two sequences that contained the same target word were not placed next to each other. The participants were asked to fill in the blank with the word they felt best completed the sequence or the word that made the sequence make the most sense. The participants were given the entire sentence with a blank in the location of the target word and not simply the context that preceded the target word; this was done because many participants in Experiment 1 were not able to press the button, and were not sure of the identity of the target word they heard, until several words after the target word had been played. Participants received and submitted their tests via e-mail. The test lasted an average of fifteen minutes.

The responses were then compiled and each context-target word pair was given a predictability index, depending on how many of the four participants were able to correctly guess the target word for the given context sequence. Thus, each sentence had an index of either 0, 0.25, 0.50, 0.75, or 1. This would suggest that for a sentence given an index of 0.25, twenty-five percent of any group of partcipants would be able to correctly guess the target word for that sentence. If subjects did not guess the exact target word but a word that was semantically related (in the case of "lawyer" for the target word "attorney"), the guess was permitted to count toward the predictability score as if it were a correct guess. Presumably, in the "lawyer" cases, if subjects of Experiment 2 were able to guess "lawyer" from the context sequences, then subjects of Experiment 1 may well have been primed for "attorney." The predictability index for each context-target word pairing can be found in the appendix.

Results

The results of Experiment 1 showed a significant effect of context style by items ($p = .047$), and a significant interaction of context style and target style by subjects ($p = .045$). The

Table 5: Mean reaction times in ms per predictability value, in both context style conditions. Note that no target word was always predictable from a certain context, and that some predictability values were not available in both conditions.

Context	Predictability				
	0	0.25	0.5	0.75	1
Full	533	478	422	n/a	n/a
Reduced	547	557	n/a	483	n/a

results of Experiment 2 build on the results of Experiment 1; the predictability index mentioned above was included as a predictor variable in a new repeated measures ANOVA of the reaction time data from Experiment 1. If reaction time difference between full and reduced contexts were affected in some way by the predictability of the target word in context, then we would expect predictability interactions in this analysis.

There was no main effect of predictability [$F_1(1,25) = 2.19$, $p = .15$] or target style [$F_1(1,26) = 1.01$, $p = .32$] and also no interaction between these variables [$F_1(1,28) = 1.46$, $p = .24$].

Contrary to the results from Experiment 1, there was a main effect of context style [$F_1(1,27) = 6.80$, $p = .01$]. This effect was not present in the first experiment.

There was a significant interaction of predictability and context style [$F_1(1,29) = 5.36$, $p = .03$]. The key finding of Experiment 1 was also found here, the significant interaction of target style and context style [$F_1(1,30) = 4.90$, $p = .03$].

There was no significant interaction of target style, context style, and predictability [$F_1(1,31) = 1.50$, $p = .23$].

The sixteen target words showed a great degree of variation in their predictability patterns. It was not the case that some of the context sentences contained words and phrases that caused all subjects to guess the missing target word; even those context-target word pairings that seemed to give away the identity of the target word only received a predictability index of 0.75. Table 5 shows the mean reaction times for each predictability index in both the full and reduced context conditions. Yet, given the high predictability scores for some of the context-target word pairings, it is surprising to see that the predictability variable itself did not show a main effect. While it did display a significant interaction with context style, it does not interact significantly with target style and context style together, and does not greatly affect the signficant interaction from Experiment 1. Predictability, then, does not seem to be a good predictor of the variance in this data. This confirms the robustness of the interaction found in Experiment 1, that even with some degree of predictability in the context sequence, the subjects' recognition of the targets was mostly slowed by the mismatch of pronunciation style between the context and the target.

The fact that the key interaction of Experiment 1 was repeated in Experiment 2 again strengthens the hypothesis that subjects are utilizing context pronunciation style to aid in lexical recognition. If the interaction can be found with the predictability data taken into consideration, the slight possibility that some subjects might be able to predict the target word from context is not the factor behind faster or slower reaction times. While a given context's predictability score may have affected some subjects' reaction times, it is clear from these results that the interaction between pronunciation style of the context and target remains the key influence on subjects' reaction times and, consequently, on lexical recognition, at least in this task.

It appears that the words that were most predictable from context were a number ("seventy") and the denomination "Republican." These words were most often directly followed by words that betrayed their identity ("seventy" was usually followed by another number, as in "(blank)-five" and "Republican" or some other partisan group was always chosen when followed by "party." See Appendix for a list of all of the context sentences and their target words). As a result of the way that the stimuli were created, it was extremely difficult to find target words in the corpus that met all of the necessary criteria (explained in the materials section of Experiment 1) and thus was very difficult to avoid situations such as these, though it would have been preferable to have all of the context sentences equalized for predictability.

General discussion and conclusion

The goal of this study was to examine phonetic reduction and the effect of preceding context in lexical recognition processes and to determine the cause of the difficulty listeners have when they hear phonetically reduced forms. I show with the results from Experiment 1 that the style of pronunciation in the preceding context before a target word plays an active role in the recognition of the target word. The mismatch in style of pronunciation between the target word and the context sequence seem to cause subjects more difficulty than simply recognizing the reduced targets in either condition. The fact that the full targets are less easily processed when embedded in a reduced context indicates that this mismatch is more important to listeners than the problems inherent in recognizing the reduced targets. A simple account of lexical access would suggest that the closer a pronunciation is to the canonical pronunciation variant of the word, the easier it will be to recognize the particular pronunciation. This study instead shows evidence that a mismatch of pronunciation style between target word and surrounding context slows auditory word recognition.

Returning to the questions posed in the background section, above, this effect sheds light on the nature and lexical status of the phonetically reduced form. If the reduced form can be processed faster than the full form when it is embedded in a reduced context, it would seem that the reduced form has as much status in the lexicon as pronunciation variants that vary in some other respect, such as vowel quality. Because they are recognized faster than the full forms in some conditions, it would seem implausible if in every instance the listener had to perform some kind of on-line repair strategy to return the form to the conventional pronunciation before recognition and access could take place. The reduced form needs, then, to be given a status in the mental lexicon that is closer to that of the conventional pronunciation than traditional architects of the mental lexicon would have suggested.

I believe that this result is line with the body of research that has shown the listener to be able to attend to very fine-tuned phonetic variation among tokens. Tokens that vary from the conventional pronunciation in any respect are able to be understood because the listener understands them in context: the idiosyncratic speech patterns of a close friend, a talker with a pronounced regional dialect, the different speech patterns of men as compared to those of women. Phonetic reduction is a very important and often ignored property of tokens of lexical items in everyday, casual speech. The degree of articulation in a dictionary-style pronunciation of a word is at a level that is not usually reached in everyday speech; being that conversational speech is a context with which listeners have much experience, just as they have experience with the speech patterns of their close friends and speakers from other dialects, it is important to investigate phonetic reduction as a variable along which each talker varies and to which the listener is very capable of paying attention. The current study offers an important look into this much ignored property of speech; many studies use single utterances elicited from a talker specifically for the purposes of the study, which are usually much more carefully articulated and carry more information than the utterances that talkers produce in everyday speech (such as words like *writer* and *button*).

One remaining question is the extent to which current models of lexical recognition are able to handle the result found here. For the model to correctly predict the behavior shown in Experiment 1, some type of context component needs to be present in the model to be able to retain information about the phonetic characteristics of an entire sequence, and make this information available to the level at which the listener is accessing and recognizing lexical forms. I do not suggest that there is an external, separate module that is solely responsible for this, simply that at some level of the architecture, the processor needs the capability of both having scope over an entire sentence or phrase and making this information available to the lower level of processing where word recognition occurs. Most models already contain this higher level of processing where syntactic and semantic properties of the words being accessed are considered in relation with the syntactic and semantic properties present in some amount of preceding context, so the models already display a level where the scope of the input being considered can be an entire sentence. Perhaps, then, style of pronunciation is computed similarly to these variables on this higher tier of processing. If, on this level, the processor can analyze a syntactic property of a word being accessed with respect to the syntactic restrictions already compiled for the sequence preceding it, it would not be difficult to conceive of a similar process with style of pronunciation.

Two of the most widely-used accounts, however, would not permit this to occur: Marslen-Wilson (1987)'s cohort model and the Shortlist model outlined in Norris (1994) do not permit information from this higher level of processing to inform the low-level word recognition process. Both authors claim that the recognition process needs to be autonomous and simply sends a constant feed of information about the word candidates to the higher level of processing, where syntactic and semantic restrictions are considered. These models would then have difficulty handling the result of the experiments detailed here, because the level on which the style of pronunciation of an entire sequence is critically informing the word recognition level and perhaps ruling out tokens that do not match in style of pronunciation. The result may be better expressed in the TRACE model (McClelland & Elman, 1986); in this account, there is constant interaction between both levels, at each recognition point. With the ability of the higher level to inform the word recognizer, contextual information could make selectional restrictions on the list of candidates competing for word recognition. However, this model has been widely criticized for this constant feed of information between the levels; Norris (1994) claims that a model such as this is implausible simply because of the cognitive burden it would imply. Despite this, it seems from the results of the current study that contextual information requires some feedback from a higher level of processing; these contextual effects should continue to be examined as they offer the investigator an opportunity to examine just how much information the listener holds in attention.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *115*(6), 3171–3183.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306-321.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*(1), 89-95.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Sjolander, K., & Beskow, J. (2000). WaveSurfer: an open source speech tool.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415-433.

Acknowledgments

This work benefited greatly from the help of many individuals. While the following individuals were consulted for their advice and technical knowledge, I take full responsibility for its implementation and any errors that may have occurred herein.

I would like to thank Ronald Sprouse and Grant McGuire for their help in designing the experiment using E-Prime and for helping me with computer-related problems in general. I am also very grateful to Klinton Bicknell and William D. Richard for their help in creating the scripts needed for the purposes of data analysis and for their statistical knowledge. The work greatly benefited from conversations with these two students and with Reiko Kataoka.

I would like to thank Dr. Larry Hyman, the secondary reader for this paper, for his help in revising and editing the drafts of this paper.

Finally, I am deeply indebted to my advisor for this project, Dr. Keith Johnson. Dr. Johnson was responsible for getting me interested in this topic, and I thank him for introducing me to the Buckeye corpus and to this research topic in particular. I am very grateful for his technical help, advice, patience, and mentorship over the course of this project.

Stimuli for Experiment 1

Below is a list of all of the context sentences with their respective target words, shown in boldface.

Experimental set

1. Twelve thousand dollars, and my **attorney** looked at him [Predictability = 0]
2. Out of that million, the **attorney** in the firm [0]
3. Um, so, my **attorney** jumped it back up [0]
4. I ended up settling; the **attorney** got most [0]
5. Nothing there that really I can **remember** that did anything [0.25]
6. Tell you the truth, I can't **remember** if they taught [0.5]
7. I don't **remember** that at all [0]
8. I don't **remember** how many months [0]
9. I would **certainly** try very hard to be [0]
10. I can **certainly** see that [0]
11. But it's **certainly**, there's a lot of things [0]
12. Things that are **probably** positive enough about it [0]
13. I think that that would be **probably** very difficult [0]
14. And then I figure I'm **probably** gonna wait tables [0.25]
15. A war that people didn't think was **important** [0]
16. The negatives in a positive way, and that's **important** [0]
17. And that's really an **important** issue right now [0]
18. Well, **usually** normal kids [0]
19. They have great ideas, but **usually** they can't [0]
20. Disability is **usually** rooted somewhere in the languages [0]
21. Well they **usually** go together; some kids have learning disabilities [0]
22. I'm an undergrad, double major in German and **history** [0.25]
23. Currently working on a senior honors thesis in **history** [0]
24. In studying **history**, I'm exposed to a number [0]
25. Rather [pause] gruesome events in **history** that still are very [0.75]
26. I also found something **fascinating**: um, I uh came [0]
27. Still on speaking terms, which is **fascinating** [0]
28. One of the **fascinating** things, when I go home [0]
29. Lock on it; put **seventy** dollars in there [0]

30. You have like maybe **seventy**-five, eighty airplanes [0.75]
31. Turn back around, and there's like **seventy**-five people waiting [0.25]
32. That's not necessarily his **opinion**, so I'm trying [0]
33. Use the word **opinion**, but I know that's not right [0]
34. That's the key in my **opinion**: personal versus interpersonal communication [0]
35. Go in the exact **opposite** direction, and I knew [0.5]
36. Or I was going in the **opposite** direction [0.25]
37. And I went the **opposite** direction; I got plugged in [0]
38. The Christianity uh, you know, like...the **opposite** of Charismatic [0]
39. Because nobody is more **similar** to you than that sibling [0]
40. Is extraordinarily **similar**, um, so what she says might [0]
41. Is often so **similar** to where I'm coming from [0.25]
42. We were able to communicate in, uh, very **similar** fashions [0]
43. Don't learn well by **listening**...maybe they don't learn [0]
44. You're explaining, he's not **listening**, he...you're trying to show him [0.25]
45. Songs that their friends are **listening** to, and their friends [0.75]
46. They're still **listening** to music that younger kids would like [0.75]
47. Walt Disney movie that doesn't have **violence** in it [0]
48. *Beauty and the Beast*, there's extreme amount of **violence** in there [0]
49. And there was lot of **violence**, and then I love [0]
50. A tremendous amount of **violence** in front of everybody's face [0]
51. It's a very nice, very small **university** but it's nice [0]
52. Mainly I suppose because of the **university** [0]
53. Or you could do the **University** of Oregon [0]
54. I don't think there's a **university** at Portland [0]
55. Cause he's the **Republican** I said I don't even want [0]
56. He's been a lifelong **Republican** uh and there's always interesting discussions [0]
57. Get the nomination because of who controls the **Republican** party [0.25]
58. During his presidency it was some **Republican** right-wing [0]

Distractor set

No response: target word presented on screen does not appear in sentence.

No-edit group: no splicing was done to the sequence

1. Need to have that established within our educational system (target: policies)
2. The Catholic church cause that was where I was raised (target: ideals)
3. We want to continue this; we want to make (target: direction)
4. Exercise your most important freedom which is to go vote (target: privilege)
5. Danced with glowsticks, and I've had I had like seven guys (target: happening)
6. Someone else's answer, they better be able to back it up (target: necessary)
7. Reelect him because he stands up for the minority (target: candidate)
8. Technology puts everything I need at my fingertips (target: amazing)

Edit group: one word (not the target) in each sequence was spliced from another location in the corpus

1. Now they live together so he's like goin' (target: apartment)
2. Uh, does not understand why people want to have, like, Izod (target: disapprove)
3. That was a wonderful thing that you did (target: benefit)
4. Everybody around the neighborhood, but I think that that's still (target: gardening)
5. Might be something as general as, 'Tell about your neighborhood' (target: description)
6. 80 IQ for example...and you know he's not (target: childhood)
7. They have beautiful music, but I...I swear I haven't seen (target: performance)
8. French is a horribly popular language for different countries to have (target: secondary)

Response : target word presented on screen does appear in sentence.

No-edit group: no splicing was done to the sequence

1. Kinds of things being **developed** on the peripheral
2. Go through when you're a **teenager**, so you're like, 'So what?'
3. That was pretty **miserable**, but it's been...t's been a while
4. Out there if somebody's **interested** in getting that kind
5. Used to like turning my **radio** on and drowning out
6. William will lose something **expensive** by the end of the year
7. One disagreement i have with intentional **communities**-that they stay open
8. Shared everything that we had, and **supposedly** we had a true

Edit group: one word (not the target) in each sequence was spliced from another location in the corpus

1. My daughters have like volunteered for the **library** reading program (splice: volunteered)

2. As much as possible; we ended up **settling** for, you know (splice: possible)
3. For each ea-each election; and sometimes I voted straight **Democrat** cause I (splice: election)
4. People who do politics of the **negative** understand what it means (splice: politics)
5. Will attack a problem a **totally** different way than the norm (splice: problem)
6. See, these people just...their brain's wired **differently** than ours is (splice: people)
7. In exchange for **postponing** it again, my rationale was (splice: exchange)
8. Most of my paintings were **completed** uh before (splice: paintings)