JOURNAL OF MEMORY AND LANGUAGE 26, 489–504 (1987)

Talkers' Signaling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction

CAROL A. FOWLER

Dartmouth College and Haskins Laboratories

AND

JONATHAN HOUSUM

Dartmouth College

An experiment examines talkers' utterances of words produced for the first time in a monolog ("new" words) or for the second time ("old" words). The finding is that talkers distinguish old words by shortening them. Two experiments show that old words are less intelligible than new words presented in isolation, but probably are not less identifiable in context. We infer that talkers may attenuate their productions of words when they can do so without sacrificing communicative efficacy. Old words can be reduced because they are repetitions of earlier presented items and because of the contextual support they receive. Two final experiments show that listeners can identify new and old words as such and that they can use information that a word is old more or less as they would use an anaphor to promote retrieval of the earlier production in its context.

Bolinger (1963, 1981) suggests that when talkers utter words that are unusual in their contexts, they lengthen them. In his example, speaking of the return trip of a person who had ridden his lawn mower cross-country, one might say, "he mowed home"; in that context, according to Bolinger, *mowed* is lengthened as compared to its duration in a sentence discussing the more usual uses of lawn mowers. Possibly, then, talkers lengthen words that have little contextual support or, more generally, that have little or no other information than their acoustic signal to specify their identity.

We thank Kristen Snow for her help in collecting and analyzing the data from several of the experiments. We also thank Carole Beal and George Wolford for their comments on an earlier version of the manuscript and George Wolford for help with the statistical analysis of data from Experiment 1. The research was supported by NICHD Grant HD 01994 to Haskins Laboratories. Reprint requests should be sent to Dr. Carol A. Fowler, Dartmouth College, Gerry Hall, Hanover, NH 03755.

Perhaps compatibly, Lieberman (1963) found a difference in intelligibility of redundant and nonredundant words presented in noise. He found that a word, for example, *nine*, that had been produced in an uninformative context ("The word that you will hear is ——") was more intelligible excised from the sentence and presented in noise than the same word originally produced in a more informative context ("A stitch in time saves ——"). Hunnicutt (1985) has partially replicated and has extended these findings.

An inference from this set of observations and findings taken together is that talkers aim to provide an acoustic signal for a word that is sufficiently informative for listeners to identify the word. If the word is probable in its context talkers may provide a reduced, acoustically less informative version of the word than if the word has a low probability or is not redundant.

Why might a talker vary his or her production of a word in this way? Two mutu-

0749-596X/87 $3.00

ally compatible reasons may be offered. One is that the reduced versions of words require less articulatory work to produce, and talkers may choose to do less work when they can get away with it without sacrificing communicative efficacy (cf. Koopmans-Van Beinum, 1980). An entirely different reason is suggested by extension of Chafe's theorizing. Chafe (1974) proposes that talkers provide information to listeners in the way that they produce words to help them distinguish "given" and "new" information in discourse.

Given information is information shared by talker and listener; but more than that, according to Chafe, it is information that the talker presumes is currently foregrounded in the listener's awareness—because it has just been mentioned or because the listener is currently looking at the thing to be named, etc. By reducing their productions of words reflecting given information, talkers thereby highlight "new" information and draw the listener's attention to it. In this theory, then, the talker deploys reductions in a systematic way to highlight the most informative words in an utterance.

Chafe's "given" information is not the same as Lieberman's or Bolinger's high-probability words. Nor is the reduction he writes of necessarily the complement of the augmentation noticed by Bolinger. Whereas Chafe writes of talkers lowering their voice pitch and destressing words conveying given information, Bolinger writes of a durational lengthening of low-probability words. Nonetheless, there is enough family resemblance across these sets of observations and findings to warrant asking whether they may not point to some interesting hypotheses concerning the talker's deployment of lengthening or reduction in speech and its consequences for the listener. Possibly, talkers attenuate their productions of a word when they can without sacrificing the word's identifiability; in order not to sacrifice identifiability, they can reduce only words whose identity is determined in part by other information available to the listener. (We will call such words, words that provide "old" information.) If this scenario is accurate, then, the talker's deployment of reductions and of more careful productions is systematic, and they can provide information to a listener that the concept named by the reduced (or augmented) word is "old" (or "new").

The experiments reported here are designed to test these hypotheses in a preliminary way.[1] They are not designed to test Chafe's or Bolinger's proposals directly, but rather to address the more general hypotheses that the foregoing summary of the literature suggests.

Experiment 1

Neither Bolinger nor Chafe provides measurements of talkers' productions of low-probability or given words. One reason why they disagree on the acoustic manifestations of augmentation or attenuation, then, may be simply that they noticed different of the acoustic consequences of reduction and augmentation. The first experiment is designed to measure talkers' productions of new and old words in speech.

For these preliminary investigations, we decided to use spontaneous speech produced by talkers in a natural, or at least a nonlaboratory, setting. This has the advantages over speech collected in the context of a controlled experiment that talkers really are attempting to communicate something to someone and that they are unaware that the way in which they are speaking will be of interest to an experimenter. The procedure has disadvantages, too. One is that the investigators have no control over the talker's use of new and old information; they must make use of what-

[1] Experiments 1–3 are replications of research by the second author performed as part of his Senior Honors project at Dartmouth College (Housum, 1986).

ever is said. A more serious problem is that the contexts in which a particular word appears as new or as old information are different. This is problematic because talkers use duration, voice pitch and amplitude for multiple purposes, not just as indices of oldness or newness. In spontaneous speech, therefore, there will be other uses of these variables that will serve as sources of random noise in the measurements. Accordingly, whereas we can be confident that talkers do use a variable systematically if we find consistent differences in its values on new and old words, we cannot be confident that talkers do not use a variable just because we find no significant effects of it in our data.

For the purposes of Experiment 1, we defined a "new" word as one produced for the first time in a passage and an "old" word, a repetition of a word spoken once before in the passage, however far back. We looked only at first and second productions in the experiment and asked whether second productions of words are shorter and lower in the fundamental frequency and amplitude of their stressed vowels than first productions. Obviously, this operational definition of "old" and "new" does not provide an entirely valid indicator of redundancy, givenness or high and low probability. That is, a word may be old because a synonym for it has been presented earlier; too, many new nouns are replaced by pronouns when they are old. However, in the passages we used there were many examples that fit our definition. The definition has the advantage of allowing us to look at productions of different tokens of the same word when they are new and old.

Methods

Materials. The major source of evidence for this experiment is a monolog from Garrison Keillor's radio program, *A Prairie Home Companion*. The monolog, titled "Sylvester Krueger's Desk," lasts 18.5 min and purports to describe Keillor's days as a fourth-grader in school in the imaginary town, Lake Woebegon (Keillor, 1985). Although the monolog is not extemporaneous, as most conversation is, it was not, apparently, read, and the speech sounds spontaneous and natural.

The passage was transcribed, and 35 pairs of words were selected for analysis. Criteria for selection were that a word occur at least twice in the passage and that, if relevant, it refer to the same object or event in both productions (so "match" referring to a "tennis match" on one occasion and as a way to light a fire on the other would be excluded). If a word occurred more than twice, just the first and second occurrences were used. Words that are chronically highly probable (*of, the*) were excluded. Also excluded were pairs of words in which one production was finally lengthened (usually because it occurred at the end of a major syntactic boundary; Cooper & Paccia-Cooper, 1980) but the other was not. Otherwise, words (including some names and some phrases, such as "Labor Day" and "ten dollar bill") were considered eligible for selection, and most eligible pairs were selected. First productions were positioned nine words from the beginning of a sentence on average and 12 words from the end; second productions were the reverse: 12 words from the beginning of a sentence on average and 9 words from the end. In an analysis of variance, the interaction between first or second production and distance from the beginning or end of a sentence was marginally significant ($F(1,34) = 3.51$, $p = .07$). However, neither new nor old words tended to fall very close to either sentence beginnings or ends. A sample paragraph from the monolog with selected items underlined appears in Appendix A.

Five additional samples of speech were taken from interviews broadcast on the MacNeil–Lehrer News hour and videotaped by a colleague for another purpose. They included separate interviews with

two congressman, two senators, and one newsperson. The shortest of these five passages contained just nine eligible word pairs. All of these were selected; in the other passages, the first nine eligible pairs were selected.

Procedure. Selected words were filtered at 10 kHz, sampled at 20 kHz, digitized, and stored on the hard disk of a computer (New England Digital Company). Three measurements were made of each word: the word's duration, the average fundamental frequency (f_0) of its lexically stressed vowel, and the peak amplitude of the same vowel.

All measurements were made from a waveform display. Duration measurements were made using visual and auditory evidence of word onset and offset. Zero crossings were identified in the waveform at locations where the word looked and sounded as if it started and ended. Measurements of f_0 and amplitude were confined to the lexically stressed vowel. (For some items, for example "Labor Day," there is more than one lexically stressed vowel; in those cases, we measured the phrasally more prominent of the two stressed vowels [in the example, /ey/ from *Labor*].) F_0 measurements were obtained by counting pitch pulses in the selected vowel, measuring the duration they spanned, and transforming the measures to Hertz values. Amplitude measures were taken from the pitch pulse in the stressed vowel with the highest amplitude; measures were in volts.

Measurements were made by the first author, but a sample of them was checked by a research assistant naive to the purposes of the experiment.[2] The sample included 20 of the 70 selected words from the Keillor passage. The 20 included 10 new words and 10 old words. These were selected randomly with the constraint that the new and old words be chosen from different pairs. Correlations between the two sets of measurements on the 20 words were .99 for duration, .94 for f_0, and .91 for amplitude.

Results and Discussion

Table 1 presents the findings on the 35 pairs of words from the Keillor monolog and below that, the five sets of nine words from the remaining passages. All comparisons reveal mean differences in the predicted direction if old words were attenuated as compared to new words. In a MANOVA with new/old as an independent varaible and (log transformed) duration, amplitude, and fundamental frequency as dependent measures, the effect of the independent variable was significant ($F(3,32) = 3.82$, $P = .02$). In univariate tests, the effect of duration ($F(1,34) = 9.83$, $p = .004$) and amplitude ($F(1,34) = 4.42$, $p = .04$) were significant; the effect of fundamental frequency was marginal ($F(1,34) = 3.20$, $p = .08$). A MANOVA performed on the data from all six talkers, with talker and new/old as independent variables, the effect of the new/old variable was significant once again ($F(3,46) = 3.27$, $p = .03$). However, in this instance, only the effect of du-

[2] We thank Kristen Snow for making these measurements.

TABLE 1
MEASUREMENTS OF OLD AND NEW WORDS FROM THE KEILLOR PASSAGE AND THE FIVE OTHER PASSAGES IN EXPERIMENT 1

	Duration		f_0		Amplitude	
	New	Old	New	Old	New	Old
Keillor	562	492	119	110	1.12	1.03
Others	436	395	135	134	1.92	1.77

Note. Measurements are in milliseconds, Hertz, and volts, respectively.

ration was significant in univariate tests ($F(1,48) = 9.28, p = .004$). All talkers had overall shorter old than new words; four of six had lower amplitude old than new words; just two of six had lower frequency old words. In the same analysis, there was a significant effect of talker on the dependent measures; however, the interaction between talker and the new/old variable did not approach significance.

In the analysis just reviewed, duration, but not f_0 or amplitude showed reliable differences depending on whether a word was being used for the first or second time. However, effects even on duration were not perfectly consistent. In the Keillor passage, 25 of the 35 words (71%) had shorter "old" than "new" words; in the remaining pairs, the direction of difference was reversed. Moreover, when shortening was observed, it varied substantially in amount from 4 ms to 414 ms. Some of this inconsistency and variability can be ascribed to the fact that the speech was spontaneous and, therefore, many sources of variability in duration were uncontrolled. However, possibly in addition, shortening may differ in amount according to some variables relevant to the old/new dimension.

One source of variability in shortening is the duration of the word when it is produced as new. Possibly, longer words generally have more room to shorten and so they may shorten more. This was the case in the Keillor monolog ($r = .46, p < .01$). A more interesting source of variation is the distance between the repetitions of a word. That is, talkers may feel free to shorten their productions of words that have just been said, but not words so far back in the conversation that listeners may not remember them (cf. Chafe, 1974). Among the 35 word pairs in the Keillor monolog, the second production followed the first by 4 words at the shortest lag and by 512 at the longest. The correlation between distance (in number of words) and shortening was exactly zero; with effects of the duration of the first production partialed out, it was .13, a nonsignificant difference in the wrong direction for the hypothesis.[3]

A final source of variability in shortening was sought in the topicality of the word pairs. Chafe (1974) proposes that talkers attenuate their productions of a word if they believe that the concept named by the word is already at the focus of listener's attention. Presumably, this would include words central to the topic of the discourse, but not the less topical words. Accordingly, we asked whether shortening would correlate positively with judged centrality of a word's meaning to the topic of its sentence or of the monolog itself.

We obtained topicality ratings in a subsidiary experiment, the methods of which are described in Appendix B. In that experiment, 10 subjects read a transcription of the Keillor monolog through and then filled out a rating sheet. On the sheet, the 35 word pairs were listed along with the page and line number in the transcription where each critical word occurred. Subjects were asked to give three ratings for each pair. They were to use a 10-point scale to rate the importance of the meaning of the word to the topic of the monolog as a whole and to rate the importance of each token of the word to the topic of its own sentence.

Only the first rating predicted shortening significantly, and that correlation was negative, contrary to prediction ($r = -.38, p = .02$). That is, words judged most important to the topic of the monolog were shortened less than less important words. Neither this correlation nor the correlation with distance is consistent with an idea that talkers only shorten words they consider to be currently at the focus of the listener's attention. Instead, the correlation with topicality suggests that talkers are least willing to shorten the most important words of the passage.

In summary, our findings so far indicate that talkers do attenuate their production of

[3] Housum (1986) did find a significant negative correlation between shortening and distance in his collection of spontaneous speech.

many "old" as compared to "new" words in discourse, the attenuation appears to take the form largely of shortening, and the shortening is least for words most central to the topic of the conversation.

We might ask why a talker would shorten old words. One answer that is likely to be correct is that attenuated productions, like casual speech more generally (Koopmans-Van Beinum, 1980; Zwicky, 1972), is easier to produce than slower, more formal productions. We will not pursue this hypothesis here. Instead, in the next two experiments, we assume that, in some sense, the talker wants to attenuate productions where possible, and we ask what allows him or her to do so.

EXPERIMENT 2

In the present experiment and the next, we consider three possible conditions that may allow talkers to attenuate their productions of words. One is that the reductions may be so slight as to leave intelligibility of the words unimpaired. This hypothesis is unlikely to be correct; if it were, then talkers presumably would attenuate their productions even of new words. A second possibility is that talkers attenuate words that have been produced before, because, in identifying a repeated word, listeners can benefit from having heard it once before in the discourse. This benefit may have two, possibly related, sources. One source is a repetition priming advantage in identification of or lexical decision to previously presented words (e.g., Fowler, Napps, & Feldman, 1985; Kempley & Morton, 1982). A second source is simply that once having figured out a word's identity, especially if it is unfamiliar (for example, a name, such as *Sylvester Krueger* in the monolog), a listener need not figure it out again based only on the acoustic signal; he or she can use the signal as a way of retrieving the word from memory. A final reason why talkers may be able to attenuate their productions of some words in a passage is that the words may be partially specified by their context. Possibly, the second productions of words are, on average, more redundant with their context than are first productions. Experiment 2 tests the first two possibilities; Experiment 3 tests the last.

Method

Subjects. Subjects were 36 students at Dartmouth College who participated for course credit. They were native speakers of English who reported normal hearing.

Materials. Two versions of a test audiotape were created; each consisted of the 35 word pairs from the Keillor monolog measured in Experiment 1. The words were excised from the monolog (using as word boundaries the zero crossings identified in Experiment 1) and were presented at a rate of 1 every 5 s. Test orders on both tapes consisted of two blocks of 35 words. One member of each of the 35 word pairs occurred once in each block. In one block, 17 items were first productions and 18 were repetitions; the other block had 17 second productions and 18 first productions. Words were differently randomized in each block. The two test tapes were complements of one another. That is, where the first tape had the first production of "antique" as its 30th trial, the second tape had the second production of "antique" in that same slot. In this way, both productions of the words of every pair appeared equally often in the first and second blocks of the tape. Eighteen subjects listened to each tape.

Procedure and design. Subjects were run in groups of two to three. They were told that they would be listening over headphones to words, names, or short phrases excised from a monolog. Their task was to identify each item if possible by writing it on the answer sheet; or if they could not identify an item, to write down as much of it as they could identify. In addition, they were to circle a number from 1 to 5 on their answer sheet expressing their confidence in their answer. A rating of 5 represented the

highest degree of confidence and 1 the lowest. There was one independent variable word history (new, old); the dependent variable was accuracy.

Results and Discussion

Subjects' responses were scored in two ways. First, answers were scored correct only if they were completely correct. (So, for example, the answer "plum" to the word "plump" received no credit for its close approximation to the target word.) In a second scoring method, answers were given scores representing the proportion of the phonemes in the stimulus string that were represented in the correct serial order in the response string. Because this scoring procedure gave exactly the same outcome in pattern and statistically as the first, we do not describe it further.

Results are given in Table 2. Subjects made more errors on old words than new words ($F(1,35) = 10.31$ $p = .003$) and more errors on words presented in the experiment for the first time than words presented for the second time ($F(1,35) = 7.10$ $p = .011$). The interaction of the variables was not significant ($F < 1$). Neither independent variable had significant effects in an analysis using items as the random factor. However, one reason for this outcome was a ceiling on performance on many items. In the condition associated with the lowest performance (that is, old words that appeared in the first block of trials), subjects achieved perfect accuracy on over half of the words (18 of 35). Of words on which some errors were made, the majority of errors in both blocks were made on repetitions (60%), and the majority of errors were made on words in the first block (72%).

TABLE 2
PERCENTAGE ERRORS AND CONFIDENCE JUDGMENTS ON NEW AND OLD WORDS AND ON THE FIRST AND SECOND BLOCKS OF THE ISOLATED-WORDS PERCEPTION TEST OF EXPERIMENT 2

Block of test	Occurrence in monolog	
	1st	2nd
	Percentage errors	
1	11.6	16.2
2	8.8	12.4
	Confidence judgments	
1	4.47	4.26
2	4.69	4.49

There is, of course, the possibility that the improvement subjects show on the second block of trials is due to a more general practice effect than the one we have been considering. That is, subjects' ability to identify words excised from context may improve with experience, and the improvement on members of word pairs that are presented second as compared to first may be a consequence of their later presentation in the test list. We looked for evidence of a practice effect of this sort by comparing performance across Trials 1–12, 13–24, and 25–35 in Block 1. Contrary to expectation, if the improvement for Block 2 items was a general practice effect rather than a specific effect of having heard other tokens of Block 2 words before, performance was nonmonotonic over the successive thirds of the first block. Performance was lowest in the middle third and slightly better in the two flanking thirds; performance in the first and last thirds was nearly identical.

Analysis of confidence judgments gave a outcome similar to the analysis of error percentages. In both the subjects and items analyses, effects of the old/new variable (subjects: $F(1,35) = 66.66$, $< .001$; items: $F(1,34) = 5.47$, $p = .02$) and of block (subjects: $F(1,35) = 32.08$, $p < .001$; items: $F(1,34) = 6.64$, $p = .01$) were both significant and consistent with the outcome on accuracy. The interaction was nonsignificant in both analyses.

Errors on old words in Experiment 2 correlated significantly with the duration difference between new and old words found in Experiment 1 ($r = .36$, $p < .05$). That is, words that had been shortened substantially in their second production were less intelligible, excised from context

and presented in isolation, than words that had been shortened less. Likewise, the more an old word had been shortened, the bigger the difference in intelligibility between the new and old word in Experiment 2 ($r = .42$, $p = .01$). Finally, the more a word had been shortened, the greater its gain in intelligibility when it was presented in the second as compared to the first block of trials (that is, when it was preceded by another production of the same word in the first block; $r = .32$, $p = .05$).

Discussion

Earlier, we proposed three possible answers to the question of what allows a talker to attenuate his or her production of an old as compared to a new word. Two of these proposed answers were addressed in the present experiment. One was that the attenuation was insufficient to affect intelligibility of the word based solely on its own acoustic signal. This was disconfirmed in the present experiment; old words were less intelligible than first productions.

The second answer was that a listener may be able to identify a word better if it follows an earlier token of the same word. This was the case in Experiment 2. Words in the second block of the experiment were identified more accurately than words in the first block. Indeed, the gain in intelligibility accruing to words in the second block was nearly enough to offset the loss of intelligibility owing to the reduction factor. That is, new words in the first block of trials were associated with an error rate of 11.6%. Old words in the second block had an error rate of 12.4%—a small difference only slightly in favor of the more careful productions. This leaves just a little work for the effects of context, examined in Experiment 3, to do for the attenuated words.

Before turning to that experiment, we should comment on a different aspect of the outcome of Experiment 2, not directly related to the questions under study. It is that words excised from context were highly intelligible in this experiment. Performance averaged about 88% correct and confidence was very high. Moreover, the performance measure almost certainly underestimates the intelligibility of words based on their own acoustic signals, but still presented in the context of the discourse. That is, the f_0 pattern, the amplitude contour, and the duration of the word reflect, in part, the word's position in its sentence and most probably its role in the sentence as well. These patterns will be at best uninterpretable when the word is presented excised from its context; at worst, they provide misleading information in their new setting. Any coarticulatory influences from neighbors likewise will present misleading information in an excised word. On the other side, Keillor was the slowest of the six talkers in Experiment 1, and he was speaking to a large audience so that the high intelligibility of his speech may overestimate that of talkers in conversation, for example (see, e.g., Pickett & Pollack, 1963).

In Experiment 3, we consider the role that context may play in facilitating identifiability of a target word.

EXPERIMENT 3

Listeners to a repetition of a word have another advantage besides having heard the word produced once before. By the time the repetition occurs,they are further into the discourse, and so they may have more information about the topic; possibly, therefore, the second occurrence of a word may generally be more redundant with its context than is the first occurrence with its context. In this experiment, we estimate that possible difference in redundancy by asking subjects to guess the target words of Experiments 1 and 2 in their contexts.

Method

Subjects. Subjects were 14 students at Dartmouth College who took part in the experiment for course credit. They were native speakers of English.

Materials. Two versions of the transcribed monolog, "Sylvester Krueger's Desk," were prepared. In each version, one member of each of the 35 word pairs from Experiments 1 and 2 was selected to serve as a test word. In one version, there were 17 new items and 18 old items. The other version was the complement of the first with 17 old items and 18 new ones. Seven subjects received each version of the monolog. The passages were printed on a computer terminal and subjects made their guesses by typing a word or words into the computer.

Procedure. Subjects were run individually. They sat in front of a computer terminal on the screen of which the monolog was gradually printed. After printing a full screen of text, the program waited for input from the subject before scrolling upward and adding more text. Thirty-five times during presentation of the text, a question mark appeared on the screen and the program stopped printing. Subjects were instructed to read the text as it appeared on the screen. When they saw a question mark, they were to try to guess the next word, name, or short phrase. They made their guesses by typing them on the terminal keyboard and hitting the return key when they were finished. The program then continued printing the text, taking up where it had left off (and therefore, providing subjects with feedback concerning their guess). Subjects were told that they could guess just one word or more than one as they wished.

Answers were scored correct if the first word typed by the subject matched the first word of the passage after the point at which the question mark had appeared. The session lasted about 30 min.

Design. The experiment had one independent variable, whether the guessed words were old or new. Subjects were crossed with the independent variable. The dependent variable was accuracy measured as the percentage of words guessed correctly.

Results and Discussion

On average, subjects guessed 18.3% of the new items correctly and 31.1% of the old items. This was a significant difference in the analysis by subjects ($t(13) = 3.79$, $p = .002$), with 13 of the 14 subjects showing effects in the predicted direction. The analysis was not significant by items, however ($t(34) = 1.54$, $p = .13$). The items analysis was nonsignificant because of a floor on performance on many items. That is, subjects made no correct guesses on either occurrence of over one-third of the test items. Of the 23 items on which at least one correct guess was made, 13 old items showed better performance than their corresponding new items, 5 new items were superior to their counterpart old items, and five pairs showed no difference at all.

Two aspects of this outcome are interesting. One is that there is a tendency for old items to be more predictable from their contexts than are new items from theirs. Perhaps more notable, however, is the finding that subjects do not very often succeed in guessing the exact next word from context. Our finding that, on average, subjects guess correctly 24.7% of the time is nearly identical to a finding by Gough, Alford, an Holley-Wilcox (1981) using a procedure in which subjects are asked to guess each successive word of a passage. This is not to say that subjects cannot often guess the *content* of the forthcoming word. Indeed, their guesses often were very close in content to that of of the forthcoming word (e.g. "old" for "antique" and "homerun" for "double"). However, the guesses infrequently corresponded to the exact next word in the passage. As Gough et al. conclude (1980), guessing from context is unlikely to play a role in ordinary reading or listening.

However, in conjunction with the rather good information for each word's identity that Experiment 2 suggests the talker provided, context can help to eliminate mishearings. (For example, it can distinguish

plum from *plump*.) Moreover, the present experiment suggests that the contextual support does tend to be better for the old words, which may require more support, than for the new items.

Returning to the question posed earlier as to what allows a talker to produce reduced versions of old words, then, we can suggest that listeners can recoup the consequent decrement in intelligibility in two ways. They can benefit from having heard the word once before and they can benefit from the more constraining context in which the old word tends to occur.

In the final experiments, we ask whether the reduction of old words not only fails to impair the intelligibility of the talker's message, but, in addition, may even provide useful information to the listener.

EXPERIMENT 4

Here we ask whether listeners can tell whether a particular utterance of a word is new or old. If they can, then possibly they can use information that a word is old to distinguish given from new information as Chafe proposes, or they can use it as a listener uses an anaphor to recover prior mention of the concept in its context (e.g., McKoon & Ratcliff, 1980).

The design of the present experiment presents listeners with a more difficult judgment than they confront in listening to continuous discourse. We presented new and old words in isolation and asked subjects to identify each as new or old. As we have pointed out before, the fundamental-frequency contour, amplitude contour, and durations of a word in part reflects its position and role in a sentence. Pulled from context, these acoustic properties of words may be more than uninformative; they may be misleading. However, if, even under these adverse conditions, listeners can make the distinction, we can be sure that they can make it in context too.

Method

Subjects. Subjects were 18 students at Dartmouth College who received course credit for their participation. They were native speakers of English who reported normal hearing.

Materials. We used the audiotapes created for the identification test of Experiment 2.

Procedure. Subjects were run in groups of one, two, or three. The experimenter explained to them that talkers attenuate their productions of words the second time they say them as compared to the first time and that the purpose of the experiment was to learn whether listeners could tell, from the way a word is spoken, whether it is being said for the first or the second time. The subjects' task was to listen to each word on the tape and to write either a 1 or a 2 on their answer sheet, 1 signifying a guess that the talker had not uttered the word before, 2 signifying a guess that the talker was saying the word for the second time. We also gave subjects the information that the tape consisted of 35 pairs of words, each pair consisting of a first and a second production of a word; therefore, on average, they should distribute their responses evenly among 1s and 2s. Nine subjects listened to one version of the tape and the remainder listened to the second version.

Results

Over the 70 trials of the experiment, subjects averaged 60% correct. Fifteen of the 18 subjects performed numerically better than chance, two subjects were at chance, and one was numerically below chance with 33 of 70 items correct. A paired t test comparing performance to the chance value of 50% was highly significant ($t(17) = 6.02$, $p = < .001$).

Although subjects found the task very difficult and made many errors, almost all of them could do the task. This lends some encouragement to an idea that listeners do have information available in the way words are pronounced in spoken discourse that indicates whether the word has been uttered before. In the final experiment of

this study, we ask whether listeners use that information in comprehending speech.

EXPERIMENT 5

In this experiment, we ask whether reductions of words when they are old actually promote comprehension by facilitating integration of related material in the discourse. That is, attenuated old words provide information in the way they are said that a word—and presumably what is being said about it—refers back to something said earlier. In this way, reduction may serve a role similar to the role of pronouns and other anaphors. A pronoun is generally less audible than the word, name, or phrase it replaces; it is short and often destressed. Despite that, it may be more informative than an exact repetition of the item it replaces because, being a pronoun, it announces its referent having been mentioned before. This may facilitate the listener's retrieving the previous relevant information and connecting it with what is being said now. Other anaphors, for example, referring to a previously mentioned car as "the vehicle" may likewise signal that the label names a previously mentioned concept.

Experiment 5 was designed to ask whether reductions work in the same way. The experiment was designed by analogy with work done on anaphoric reference by McKoon and Ratcliff (1980). In two experiments, they showed that anaphors activate not only the words they replace, but also other words from the same proposition as the words they replace.

A sample paragraph from their experiments follows: "A burglar surveyed the garage set back from the street. Several milk bottles were piled at the curb. The banker and her husband were on vacation. The burglar/the criminal/a cat/ slipped away from the streetlamp."

In the experiment, different groups of subjects saw the three versions of the last sentence. In one version, a word from the first sentence reappeared in the last sentence: in another, an anaphor appeared, and in a third, an unrelated word appeared. After reading a paragraph, subjects were given a test word on which to make an "old" or "new" response depending on whether the word had appeared or not in the preceding paragraph.

McKoon and Ratcliff found that anaphors (e.g., "the criminal") in the last sentence of the paragraph were as effective as repetitions of previously mentioned words (e.g., "the burglar") in activating the test word (e.g., "garage") from the first sentence of the paragraph; both versions of the last sentence led to a faster "old" decisions to the test word than did the version with an unrelated subject noun phrase ("a cat").

In a second paradigm, subjects read two paragraphs and then made old/new decisions to a list of test words. In that experiment, a word from the last sentence (e.g., "streetlamp") was judged more rapidly preceded by "burglar" than preceded by a word from the other paragraph. This outcome occurred both for subjects who had seen "burglar" in the last sentence and for subjects who had seen "criminal."

Although priming differences between the original word and the anaphor were very small and nonsignificant, numerical differences favored the anaphor in both experiments.

In the present experiment, we used a procedure similar to those of McKoon and Ratcliff to ask whether a reduced version of a spoken word might serve as a *better* reminder of words in a sentence containing a nonreduced version of the same word than would the nonreduced version itself.

Method

Subjects. Subjects were 33 students at Dartmouth College who received course credit for their participation. They were native speakers of English who reported normal hearing. Data from two subjects were eliminated because of poor performance (near chance accuracy in one instance, response times averaging twice

those of the remaining subjects in the other).

Materials. Once again, the Keillor monolog was used. From the monolog, 42 prime target pairs were selected on which subjects would make judgments whether or not the word had occurred before in the passage. (These were called "old"/"new" decisions in speaking to subjects; however, to avoid ambiguity with another use of "old" and "new" in this manuscript, we will call them "yes"/"no" recognition decisions.) At 42 selected locations in a re-recording of the monolog, a 1000-Hz tone was placed on the other channel of the tape than the channel used to present the monolog to subjects. This tone, input to a computer, caused the program running the experiment to stop the tape recorder and to present a warning tone followed by two words on which subjects made speeded recognition judgments. We will call the first word of each pair the "prime" and the second word the "target" for that trial. Response times were measured from prime and target onset.

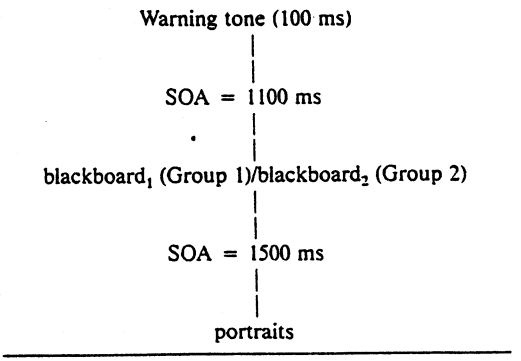
Of the 42 prime-target pairs presented to subjects, 14 were critical pairs and the remainder were fillers. In the critical prime-target pairs, both words had occurred recently (12 syllables back on average) in the monolog and so the correct recognition judgment was "yes." In all of the critical pairs, the prime was one of the 70 words measured in Experiment 1 and further studied in Experiments 2–4. In seven pairs, the prime was the same version of the word produced recently in the passage and it was the first occurrence of the word in the monolog. In seven pairs, the prime was not the same version of the word as that just produced in the monolog and it was the old version of the word. On these trials, the target was some other word, near the nonreduced prime and in the same sentence as the nonreduced prime in the monolog. The 14 critical primes were selected based on their distribution throughout the monolog; they were not selected based on the durational difference between new and old versions of the primes. Table 3 shows how a trial was organized in the experiment.

Two versions of the experiment were run, one on 16 subjects and the other on the remaining 15. The versions were complementary so that if subjects in the first group had the new version of a word as a critical prime, subjects in the second group had the old version on the same trial.

Filler trials included six trials in which both prime and target had not occurred in the monolog at the point where they were tested. (That is, the correct response to both prime and target was "no.") In this and other filler trials, words on which a "no" response was correct were selected from the monolog but from a location further on than the point where the words were tested. Eleven trials each were "yes"–"no" and "no"–"yes" trials. Trials of the various types occurred in quasi random order and occurred at irregular intervals throughout the monolog. The first critical trial was the fourth trial of the experiment.

Design. There was one independent variable, whether the critical prime was a first or a second production. Dependent variables were response time to the target and accuracy of response to the target. In addi-

TABLE 3
SAMPLE TRIAL FROM EXPERIMENT 5
Oh that smell about it, same blackboard, same portraits of Washington and Lincoln up front and center



tion, we looked at response times and accuracy to primes depending on whether they were first or second productions.

Results and Discussion

Response times to targets were included in the analysis only if the response was accurate and if the response to the prime had been accurate. There were no correct response times to critical primes or targets slower than 2500 ms; no responses were deleted from the analysis because of their duration.

Table 4 presents the mean response times and proportions correct for critical primes and targets. On targets, the accuracy measure reflects the number of correct responses independent of accuracy on primes.

Responses to reduced primes were overall faster than to first productions. The difference was significant in the subjects analysis only ($F(1,29) = 11.31$, $p = .002$). The same analysis showed no effect of group and no interaction of group × prime type. The difference between new and old primes was not significant in an analysis by items ($t(13) = 1.32$, $p = .21$). Of the 14 critical items, nine showed a difference favoring the reduced prime.

The difference in response times to the new and old primes, significant in the analysis by subjects may, in any case, reflect only the duration difference between reduced and unreduced words. This difference averaged 89 ms for the 14 critical items of the experiment.

TABLE 4
AVERAGE RESPONSE TIMES AND PROPORTIONS OF CORRECT RESPONSES TO NEW AND OLD PRIMES AND TO TARGETS PRECEDED BY NEW AND OLD PRIMES, DATA FROM EXPERIMENT 5

	Prime		Target	
	New prime	Old prime	New prime	Old prime
RT	834	793	758	719
Accuracy	.92	.90	.89	.91

Response times to targets are faster following old primes than following new primes. Results are weak but significant in both subjects and items analyses (subjects: $F(1,29) = 4.15$, $p = .048$; items: $t(13) = 2.25$, $p = .04$). The small accuracy difference also favors targets preceded by old primes; however, the difference did not approach significance in either analysis by subjects or by items.

The significant difference in reaction time apparently cannot be explained simply as faster response times to targets that follow short primes or that follow fast responses to primes. Correlations between response times to targets and prime durations, and between response times to targets and response times to primes (computed separately on new and old primes to eliminate effects of the independent variable) are uniformly nonsignificant.

In conjunction with Experiment 4, the present experiment shows both that listeners can distinguish reduced from unreduced versions of a word and that they can use the perceived reduction as information that a word has been mentioned before to facilitate recall of the word's prior context.

GENERAL DISCUSSION

We have found that talkers attenuate their productions of old words and that the identifiability of these redundant words is affected if the words are presented in isolation, but is probably not affected for the words in context. Finally, we have found that listeners can identify words as old or new, and they can use information that a word is old to facilitate integration of related material in a discourse.

If talkers reduce old words, as we suppose, for "selfish" reasons—an idea, it is true, that requires experimental test—then the present study reveals an interesting example of a sort of "symbiotic" relationship between talkers and listeners.

Talkers may reduce their productions of old words because it is easier to produce reduced than careful versions of words,

and because listeners do not need as good a signal for an old as for a new word. Listeners do not need as good a signal because, having heard a word before, they find it relatively easy to identify it a second time, and because the context of an old word tends to be more constraining than that of a new word. By reducing old as compared to new words, however, talkers deploy reduction systematically and therefore, reduction (or on the other side, careful articulations) can provide information to a listener that a word relates back to something said earlier (or does not). As Experiment 4 shows, listeners can tell reduced from unreduced words even under quite adverse conditions in which the words are excised from their context. Experiment 5 shows that they can take advantage of the information provided by reductions to retrieve the earlier context of the word.

Possibly, this instance of a behavioral systematicity that is beneficial for different reasons both to talkers and to listeners is not unique to production and perception of new and old words in speech. Indeed, possibly this confluence of mutual benefits may promote the prepetuation of various systematic behaviors in a language and across languages.

That is, there may be other examples in which talkers produce speech in certain ways because it is easier to than not, but, given that they do, the listener is provided with useful information. One possible other example is declination—the tendency for the fundamental frequency of the voice to drift downward over the course of a coherent syntactic unit (e.g., Cooper & Sorenson, 1981). Other things equal, f_0 will decline during an expiration as the lungs deflate. Declination due to this effect is observed even in word sequences produced with no communicative intent (Sternberg, Wright, Knoll, & Monsell, 1980). Talkers tend to take breaths at major syntactic (or metrical) boundaries (e.g., Grosjean & Collins, 1979) so, other things equal, f_0 will rise there too.

Therefore, declination and resetting will tend to be deployed systematically even though the talker is essentially just letting declination happen during expiration. Because f_0 resetting is systematic, however, the listener can use it as redundant information demarcating major syntactic boundaries.

Of course, the whole account of declination may be more complicated (see, for example, Cooper & Sorenson, 1981, who think that it is much, much more complicated). It has been highly controversial whether declination can be seen as an automatic consequence of lung deflation or instead must be seen as an intentional imposition by the talker (compare Cohen, Collier, and t'Hart, 1982; Cooper & Sorenson, 1981; Gelfer, Harris, Collier, & Baer, 1983). We guess that the near universality of declination across languages (see the review by Cooper & Sorenson) is explained by the observation that it is easier for the talker to exhibit declination on expiration than not. However, because declination is informative, and because listeners use the information (see, e.g., Breckinridge, 1977), talkers may on occasion use declination and resetting intentionally to provide information at a boundary at which the talker does not need to take a breath.

More generally, we hypothesize that the confluence of articulatory ease and perceptual redundancy may promote perpetuation of systematic deployment of various kinds of articulatory information in speech.

Appendix A

Excerpt from "Sylvester Krueger's Desk" by Garrison Keillor

Oh that smell about it, same blackboard, same portraits of Washington and Lincoln up front and center, up over the blackboard, Washington on the left, Lincoln on the right. Looking down on us all these

years like an old married couple up there on the wall. I'd sit there at my desk, you know, bent over the paper trying to make big fat vowels so that the tops of them would just scrape the little dotted line. Make the tails of the consonants, the ps and the qs and the gs and fs so that they hung down. There I'd sit and memorize arithmetic tables and memorize state capitols and major exports of many lands. And whenever I was stumped, I'd always look up to see their pictures.

APPENDIX B

Methods for Topicality Rating Study

Subjects. Subjects were 10 students at Dartmouth College who participated for course credit. They were native speakers of English.

Materials. Subjects were given a typed transcription of the monologue, Sylvester Krueger's Desk. In addition, they received a rating sheet. On the sheet the 35 word pairs measured in Experiment 1 were listed. Next to each word were listed the pages and line numbers of its first two occurrences. In addition, there were slots for three topicality ratings.

Procedure. Subjects were run in groups of two to four. On arrival they were given copies of the seven-page transcription and they were asked to read it through quickly. As each subject finished, he or she was given a rating sheet and typed instructions. The instructions asked subjects to locate each relevant occurrence of a word in the passage and to rate the word's importance to the topic of its sentence. Next they were to rate the importance of the word's meaning to the topic of the monologue as a whole. Examples were provided from a different text to illustrate important and less important words in their respective sentences and passages.

Subjects reported no difficulty in following instructions. The session lasted about one half hour.

REFERENCES

BOLINGER, D. (1963). Length, vowel, juncture. *Linguistics*, 1, 5–29.

BOLINGER, D. (1981). *Two kinds of vowels, two kinds of rhythm*. Bloomington, IN: Indiana University Linguistics Club.

BRECKINRIDGE, J. (1977). Declination as a phonological process. Murray Hill, NJ: Bell Laboratories Tech. Memo.

CHAFE, W. (1974). Language and consciousness. *Language*, 50, 111–133.

COHEN, A., COLLIER, R., & T'HART, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, 39, 254–273.

COOPER, W. E., & PACCIA-COOPER, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard Univ. Press.

COOPER, W. E., & SORENSON, J. (1981). *Fundamental frequency in speech production*. New York: Springer-Verlag.

FOWLER, C. A., NAPPS, S., & FELDMAN, L. (1985). Relations among regular and irregular morphologically-related words in the lexicon as revealed by repetition priming. *Memory & Cognition*, 13, 241–255.

GELFER, C., HARRIS, K., COLLIER, R., & BAER, T. (1983). Speculations on the control of fundamental frequency declination. *Haskins Laboratories Status Reports on Speech Research, SR-76*, 51–63.

GOUGH, P., ALFORD, A., & HOLLEY-WILCOX, P. (1981). Words and context. In O. Tzeng & H. Singer (Eds.), *Perception of print*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1981, pp. 85–102.

GROSJEAN, F., & COLLINS, M. (1979). Breathing, pausing and reading. *Phonetica*, 36, 98–114.

HOUSUM, J. (1987). *Specification of given and new information in conversation*. Unpublished manuscript.

HUNNICUTT, S. (1985). Intelligibility versus redundancy—Conditions of dependency. *Language and Speech*, 28, 45–56.

KEILLOR, G. (1985). Sylvester Krueger's desk. Gospel Birds cassette tape, Minnesota Public Radio.

KEMPLEY, S., & MORTON, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, 73, 441–454.

KOOPMANS-VAN BEINUM, F. J. (1980). *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers B. V.

LIEBERMAN, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172–187.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668–682.

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, 6, 151–164.

Sternberg, S., Wright, C., Knoll, R., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Erlbaum.

Zwicky, A. (1972). On casual speech. In P. Peranteau, J. Levi, & G. Pharse (Eds.), *Papers from the eighth regional meeting of the Chicago Linguistics Society* (pp. 607–615). Chicago: Chicago Linguistics Society.

(Received March 11, 1987)
(Revision received April 17, 1987)