Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 5, 1217–1232

0278-7393/97/$3.00

Word-Frequency Effects on Short-Term Memory Tasks: Evidence for a Redintegration Process in Immediate Serial Recall

Charles Hulme and Steven Roodenrys
University of York

Richard Schweickert
Purdue University

Gordon D. A. Brown
University of Warwick

Sarah Martin and George Stuart
University of York

Four experiments investigated the mechanisms responsible for the advantage enjoyed by high-frequency words in short-term memory tasks. Experiment 1 demonstrated effects of word frequency on memory span that were independent of differences in speech rate. Experiments 2 and 3 showed that word frequency has an increasing effect on serial recall across serial positions, but Experiment 4 showed that this effect was abolished for backward recall. A model that includes a redintegration process that operates to "clean up" decayed short-term memory traces is proposed, and the multinomial processing tree model described by R. Schweickert (1993) is used to provide a quantitative fit to data from Experiments 2, 3, and 4.

It is now clearly established that verbal short-term memory makes heavy use of some form of phonological or articulatory coding (see Baddeley, 1986; Penney, 1989; for reviews). According to trace decay models, verbal short-term memory is limited in capacity, with items being represented by traces that decay within a short period of time (e.g., Baddeley, 1986; Schweickert & Boruff, 1986). In such trace decay models, decay can be overcome by rehearsal that involves some form of subvocal articulation.

A simple and influential model of this type is embodied in the notion of an articulatory loop (e.g., Baddeley & Hitch, 1974; Baddeley, Lewis, & Vallar, 1984). This model provides a parsimonious explanation for many short-term memory effects. So, for example, the word-length effect (Baddeley, Thomson, & Buchanan, 1975), the fact that people can recall more short than long words in order, has been explained in terms of long words being spoken and so rehearsed more slowly than short words. Phonologically similar words are recalled less well than dissimilar words, and this has been attributed to the detrimental effect of similarity between traces on retrieval from a phonological store (for reviews of these effects, see Baddeley, 1986; Hulme & Mackenzie, 1992).

One variable that cannot be accommodated in a straightforward way by such trace decay models is word frequency. Word frequency has been shown to affect both long- and short-term memory tasks. In long-term memory tasks, performance is generally better on high-frequency words than on low-frequency words; the notable exception to this pattern being the episodic recognition task (e.g., Mandler, Goodman, & Wilkes-Gibbs, 1982). Such episodic recognition tasks require access to knowledge about a particular encounter with a given word; this, however, is a very different task from the short-term immediate serial-recall tasks we are concerned with in the present article. In these tasks participants appear to maintain an ordered representation of the phonological forms of words that have been presented on a given trial. We argue below that the processes involved in short-term memory tasks are sensitive to the properties of the representations of items held in lexical memory and that frequency has effects on recall through its effects on these lexical representations. High-frequency words are recognized more easily in speech than are low-frequency words (e.g., Howes, 1957), and our arugment is that processes analogous to speech perception are involved in short-term recall tasks.

In short-term memory tasks the effects of frequency are complex and not well understood. Watkins (1977) found that memory span scores (the mean number of words that could be correctly recalled in order) were higher when the first half of the list comprised high-frequency words and the second half comprised low-frequency, than when this arrangement was reversed. He argued from this that there was a long-term memory component in the span task. The greater effect of frequency in the initial part of the list was interpreted in terms of items early in the sequence being recalled from

Charles Hulme, Steven Roodenrys, Sarah Martin, and George Stuart, Department of Psychology, University of York, York, United Kingdom; Richard Schweickert, Department of Psychology, Purdue University; Gordon D. A. Brown, Department of Psychology, University of Warwick, Warwick, United Kingdom. Steven Roodenrys is now at the Department of Psychology, University of Wollongong, Wollongong, New South Wales, Australia.

This research was supported by Grants R000 232 567 and R000 236 216 from the Economic and Social Research Council of the United Kingdom and also in part by National Science Foundation Grant 9123865-DBS. We thank Dave Balota, Richard Chechile, Ian Neath, and Philip Quinlan for helpful discussions, and Xiangen Hu for help with his parameter estimation program. We also thank John Wixted for his constructive comments on the article.

Correspondence concerning this article should be addressed to Charles Hulme, Department of Psychology, University of York, York YO1 5DD, United Kingdom. Electronic mail may be sent via Internet to CH1@york.ac.uk.

long-term memory, whereas later items were retrieved from short-term memory. However, this experiment also showed an effect on span of the frequency of the second half of the list.

Wright (1979) showed that low-frequency words take longer to say than high-frequency words, even when they were equated for the number of letters. This difference in spoken duration can explain Watkins (1977) results because the longer duration of low-frequency words will allow greater degradation or fewer rehearsals before recall. Thus, when the first part of a list consists of low-frequency words, these words are spoken more slowly and lead to later items being subject to more degradation than in the case of high-frequency words that are spoken more quickly.

Articulatory suppression (repeating some irrelevant speech) is generally assumed to prevent the encoding of visually presented material onto the articulatory loop (Baddeley, 1986). Both Tehan and Humphreys (1988) and Gregg, Freedman, and Smith (1989) showed that although speech rate and memory span were higher for high-frequency words than for low-frequency words, the difference in memory span remained between these classes of words under articulatory suppression. Roodenrys, Hulme, Alban, Ellis, and Brown (1994) also found that word frequency had effects on short-term memory span that could not be accounted for by differences in speech rate. These findings show that at least part of the effect of word frequency on span is due to processes other than those involved in the articulatory loop. An obvious possibility is a contribution from long-term memory (Craik, 1968; Watkins, 1977).

Some recent studies have pursued the idea that one component of memory span reflects the availability of representations of the phonological form of words in long-term memory. Hulme, Maughan, and Brown (1991) compared memory span for familiar words with span for nonwords that were initially completely unfamiliar to their participants. It was found that memory span for nonwords was lower than for words, though both types of items showed an equivalent relationship between speech rate and memory span, with longer items being spoken more slowly and recalled more poorly. The difference in memory span between the words and nonwords was not attributable to differences in speech rate, however, because the nonwords were recalled much more poorly than would be expected from the speed at which participants could articulate them.

We argued that the strong relationship between articulation rate and memory span observed for both words and nonwords reflected a speech-based, time-limited component that reflects the speed with which information in short-term memory can be refreshed. We suggested that the greater difficulty of recalling nonwords as compared with words reflected the absence of a stored representation of the phonological form of the nonwords (a representation in long-term memory). According to this argument, when attempting to recall items in a serial-recall task, participants may often be dependent on a partially decayed trace of the phonological form of the words in short-term memory. However, knowledge of the spoken form of words may help the person perform *pattern completion* on decayed traces and so successfully recall them. Our suggestion was not that this involves deliberate guessing but rather that people perform such a process automatically by using processes that are normally an integral part of speech perception and speech production mechanisms. Recently, we have shown that inducing familiarity with nonwords, by having participants repeat them aloud, results in improvements in memory span for those items that appear to be attributable to participants acquiring knowledge of their phonological form (Hulme, Roodenrys, Brown, & Mercer, 1995). We have also described simulations that demonstrate that a redintegration or pattern completion process of this kind can indeed account for the improved memory performance for nonwords (Brown & Hulme, 1995).

Experiment 1

The aim of the first experiment was to explore whether the difference in memory span for high- and low-frequency words reflects differences that are analogous to those between words and nonwords in our earlier study (Hulme et al., 1991). We therefore related speech rate to memory span for high- and low-frequency words of different spoken lengths. Following the logic and findings of our previous study, we expected memory span for high-frequency words to be greater than that for low-frequency words. We also expected that these differences would be at least partially independent of any differences in speech rate between the high- and low-frequency words.

Method

Participants. The participants were 18 undergraduate and postgraduate students (11 women and 7 men) at the University of York who were paid for their participation.

Materials. Six sets of eight words were used throughout the experiment. The three high-frequency sets, of short, medium, and long duration, comprised words with a frequency of 103 words per million or more, according to the norms of Kučera and Francis (1967). The three low-frequency sets comprised words with a frequency of 5 words per million or less. On the basis of some pilot studies, we made an attempt to match the high- and low-frequency words of each length for spoken duration. All six sets were matched for rated concreteness by using the Oxford Psycholinguistic Database (Quinlan, 1992). The words used are listed in Appendix A, and their frequencies and concreteness ratings are listed in Table 1.

Procedure. The experiment was conducted in a single session lasting approximately 40 min. All of the tasks in the experiment were controlled by a Macintosh SE/30 computer that used an external amplified speaker to present the previously recorded and digitized items (see Cox, Hulme, & Brown, 1992). Each participant completed a memory span procedure and speech rate measurement for each of the six conditions. The order of testing for both variables, high versus low frequency and length, was counterbalanced across participants.

In each condition, the participants initially listened to and repeated each word to check its audibility. For the low-frequency words, we told them that there might be some words they were unfamiliar with and to indicate these. When a participant did not know a word (which occurred, on average, for less than one word per participant), we gave a definition of the word.

Table 1
Frequencies and Concreteness Ratings for Stimulus Sets Used in Experiment 1

	High-frequency words			Low-frequency words		
Variable	Short	Medium	Long	Short	Medium	Long
Frequency						
M	176.0	177.8	176.1	1.3	1.8	1.9
SD	87.0	92.2	52.0	1.8	2.0	1.7
Concreteness						
M	399.6	381.8	384.4	407.9	398.9	374.6
SD	55.8	49.5	82.3	56.7	61.0	67.7

To measure memory span, we presented participants with lists of words drawn randomly, without replacement from the pool of eight items, at a rate of one item per second. The participants were presented with two lists of items at each sequence length, beginning with three items. All participants successfully recalled both of the initial lists in all conditions. The length of the lists was increased by one item each time until the participant made errors on both lists at a given length. Memory span was calculated as the greatest list length at which the participant could recall all lists correctly, plus half a point for each subsequent list recalled correctly.

After the memory span tasks, the participants' speech rates were measured. For each condition, the participants were presented with the eight items from that condition, in four pairs. The participants were instructed to repeat each pair 10 times, as quickly as possible, until told to stop; The time taken to do this was recorded. The mean of these four times was then transformed to items articulated per second.

Results

In Figure 1, the mean memory span scores for high- and low-frequency words are plotted as a function of speech rate. As expected, it is apparent that increased item length has a detrimental effect on memory span and on speech rate. As we predicted, memory span is greater for the high-frequency words than for the low-frequency words, and this advantage does not appear to be attributable to differences in speech rate.

The memory span scores were subjected to a two-way analysis of variance (ANOVA), with repeated measures on both variables, frequency, and item length. This analysis revealed significant effects of frequency, $F(1, 17) = 48.09$, $MSE = 0.20$, $p < .001$, and of word length, $F(2, 34) = 51.09$, $MSE = 0.27$, $p < .001$, but no interaction between these two variables, $F(2, 34) = 0.01$, $MSE = 0.16$, *ns*. Planned comparisons revealed a significant difference between span scores for medium-length and long items, $F(1, 17) = 83.12$, $MSE = 0.08$, $p < .001$; and between short- and medium-length items, $F(1, 17) = 6.49$, $MSE = 0.18$, $p < .05$. Neither of these comparisons interacted significantly with frequency, both $Fs(1, 17) < 1$.

A two-way ANOVA with repeated measures on both variables, frequency and item length, was also carried out on the speech rate data. This analysis revealed significant effects of frequency, $F(1, 17) = 37.55$, $MSE = 0.04$, $p < .001$, and of word length, $F(2, 34) = 585.41$, $MSE = 0.05$, $p < .001$, but no interaction between these two variables, $F(2, 34) < 1$, $MSE = 0.02$. Planned comparisons revealed a significant difference between speech rates for medium-length and long items, $F(1, 17) = 679.9$, $MSE = 0.001$, $p < .001$, and between short- and medium-length items, $F(1, 17) = 275.54$, $MSE = 0.64$, $p < .001$. Neither of these comparisons interacted significantly with frequency.

Inspection of Figure 1 shows that frequency has an effect on memory span over and above the effect that it has by influencing speech rate. To examine this effect, we subjected the memory span scores to an analysis of covariance (ANCOVA), with speech rate as a covariate. This analysis allowed us to control for the effects of speech rate on memory span scores. If differences in memory span for high- and low-frequency words remained after the effects of

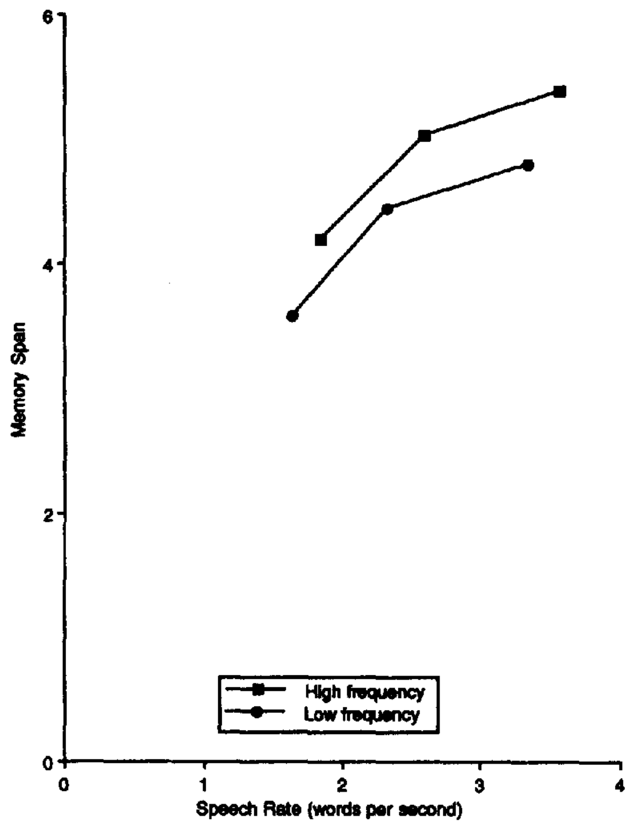

Figure 1. Memory span as a function of speech rate for the short, medium, and long words of high and low frequency in Experiment 1.

speech rate had been controlled for, we would have good evidence that the effects of word frequency are not attributable to differences in speech rate.

The ANCOVA revealed a significant effect of word frequency on memory span, $F(1, 16) = 11.56$, $MSE = 0.21$, $p < .01$, and a markedly weaker effect of word length, $F(2, 33) = 5.40$, $MSE = 0.25$, $p < .01$. From Figure 1 it can be seen that there is a quadratic component to the relationship between speech rate and memory span, thus the inclusion of a quadratic term in the ANCOVA may be warranted to remove further effects of speech rate. An ANCOVA was therefore performed on the memory span scores with two covariates, the simple speech rate measure and the square of this measure. The results showed a significant effect of frequency, $F(1, 15) = 12.09$, $MSE = 0.20$, $p < .01$, but no significant effect of item length, $F(2, 32) = 2.57$, $MSE = 0.26$, $p > .05$. Thus, adding the quadratic term to the ANCOVA removed the effect of item length, but a frequency effect remained.

Discussion

We set out to examine the source of word-frequency effects on memory span. Previous research had indicated that some effects of word frequency on span were attributable to differences in the rate at which high- and low-frequency words can be articulated (Wright, 1979). Our results have shown that word frequency, although it may affect speech rate, has effects on memory span that are independent of speech rate (cf. Roodenrys et al., 1994). However, word length effects on memory span seem well explained in terms of speech rate differences. These conclusions follow from the ANCOVA presented above. Controlling span scores for variation in speech rate abolished the effects of word length but left a highly significant effect of word frequency.

In terms of the approach outlined earlier, we have found support for the view that the difference in memory span between high- and low-frequency words (like the difference between words and nonwords) reflects some non-articulatory contribution to memory span. We believe that this processing component of memory span may reflect differences in the accessibility of the phonological representations of the items, with the representations of high-frequency words (in long-term memory) providing more effective support at retrieval for redintegrative processing operating to "clean up" degraded memory traces.

The finding that short-term memory span is better for words than for nonwords (Crowder, 1978; Hulme et al., 1991) also suggests that the availability of representations of the items influences memory span. There are at least two ways in which having a long-term memory representation of an item might facilitate performance in a short-term memory span task, and either would allow word frequency to produce the effect we have observed. The first proposal (Watkins, 1977), which was mentioned earlier, is that people recall some of the early items in a list directly from long-term memory. The long-term memory information in this case is episodic in nature, in that it requires the person to remember a specific episode (the list of items heard a few moments before).

A second alternative is that frequency, by means of its effects on the representations held in permanent memory, influences a process that operates to help identify decayed traces from a short-term store. Such a redintegration process has been postulated by a number of researchers (e.g., Brown & Hulme, 1995; Cowan, 1992; Hulme et al., 1991; Nairne, 1990; Schweickert, 1993). Our own view of this process has been to suggest that it may be useful to think of short-term memory tasks drawing on speech perception and speech production mechanisms. According to this view, automatic pattern completion processes operate to clean up decayed traces, possibly as a by-product of mechanisms that exist for the perception and production of speech. A connectionist model of such an architecture has been developed by Brown (1990). This model depends on long-term memory mechanisms, but only in the sense that it depends on knowledge of the phonological forms of words held in memory. However, it is quite different from a more traditional view that postulates separate short- and long-term stores, each of which store certain items in a short-term memory task. Brown and Hulme (1995) described a more detailed computational model of verbal short-term memory according to which redintegration is responsible for several short-term memory phenomena.

Recently, a formal model of this sort of two-component view of memory span performance has been developed by Schweickert (1993). Schweickert proposed a multinomial processing tree model (which is described in more detail later) as a way of capturing the processes that operate in recall of a random list of words. According to the model, some items can be recalled directly from undegraded phonological traces. However, if a trace is degraded somewhat, the probability that it can be reconstructed depends on whether a long-term memory representation of the item exists. Schweickert noted that "this is a long-term memory contribution, but only in the sense that it is done through knowledge of the language" (p. 170). Schweickert's model has the advantage of making some quantitative predictions about expected patterns of serial-recall performance.

Experiment 2

The aim of the second experiment was to examine the effect of word frequency on short-term serial recall as a function of serial position. This should allow us to discriminate between the two alternative mechanisms proposed for the long-term memory contribution to span tasks outlined above. If it is the case that items from early positions in the list are retrieved from long-term memory, whereas later items are retrieved from short-term memory, then we can expect to observe an effect of word frequency on early positions but not on later positions. In view of Wright's (1979) suggestion that uncontrolled differences in speech rate might account for the effects of word frequency on memory span found by Watkins (1977), we also wanted to control for differences in speech rate between our high- and low-frequency words.

Schweickert's (1993) multinomial processing tree model of immediate serial recall is illustrated in Figure 2. In a multinomial processing tree model each mental process is represented by a branch in the tree and is carried out with a certain probability. The terminal nodes of the tree correspond to observable outcomes: in this case, the correct or incorrect recall of an item. The probability that a particular terminal node occurs is the product of the probabilities of the branches in the path from the root of the tree to the terminal node. Recent work has led to convenient techniques for parameter estimation in such models that have been applied to several memory phenomena (Batchelder & Riefer, 1986, 1990; Chechile, 1987; Chechile & Meyer, 1976; Hu & Batchelder, 1992; Riefer & Batchelder, 1988).

In Schweickert's (1993) model, one process attempts to retrieve items directly from the short-term store, and a second redintegration process operates on degraded traces retrieved from this store. The probability of direct retrieval from the short-term store is labelled I in Figure 2. The probability that the second redintegration process is required is the probability that direct retrieval is not successful, $1 - I$.

There are two mechanisms in the model through which frequency would be predicted to have an increasing effect across serial positions: an effect on degradation produced by a change in speech rate and an effect on the redintegration process. To deal with speech rate first, in this model, recall of early items is assumed to be better than for later ones because early items have a higher probability of being retrieved directly from short-term memory. In relation to Figure 2, the parameter I is assumed to decrease across serial position. This is due to the increased effect of trace degradation across serial positions. The amount of trace degradation is assumed to be a function of the elapsed time since the presentation of the list (Schweickert, 1993), and because longer words will be spoken and so recalled more slowly than shorter words, we expect a greater decrease in parameter I across serial positions for lists of long words. From this, the model predicts an interaction between word length and serial position. Cowan et al. (1992) reported just such an interaction: Performance in a serial-recall task was worse with the words that took longer to say, and the difference between the sets increased with serial position. Thus, we might predict that a manipulation of word frequency will produce an interaction with serial position if, as has sometimes been found, frequency influences the pronunciation rate of the items (Wright, 1979). However, it is clear from Figure 1, that although the differences in speech rate between the high-frequency words and low-frequency words were significant, they were quite small. Differences in speech rate of this magnitude may not be sufficient to make the predicted interaction manifest.

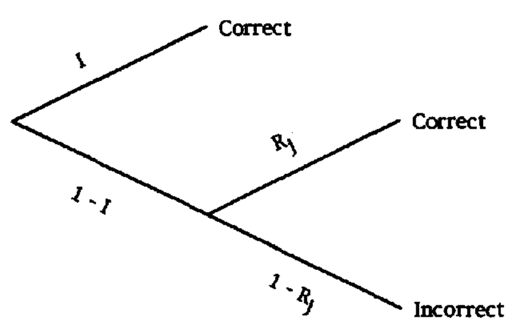

Figure 2. Multinomial processing tree model of verbal short-term memory retrieval. I = probability that an intact representation of an item exists and that an item is recalled directly and correctly; R_j = probability that a degraded item of frequency level j (where $j = 1$ for high-frequency words and $j = 2$ for low-frequency words) is correctly redintegrated. From "A Multinomial Processing Tree Model for Degradation and Redintegration in Immediate Recall," by R. Schweickert, 1993, *Memory & Cognition*, p. 169. Copyright 1993 by the Psychonomic Society. Adapted with permission.

Whether or not the first mechanism is effective, the model in Figure 2 also predicts an increased effect of frequency across serial position through the redintegration process. Later items in a list are subject to more degradation, possibly as a function of output delays (Cowan, 1992), and so recall of these items is more dependent on redintegration with knowledge from long-term memory. That is, the probability of correctly retrieving the item from the short-term store (parameter I) decreases across serial positions. Thus, the relative contribution of a redintegration process operating after failure to retrieve an item from the short-term store (parameter R) will increase across serial positions.

We assume that this redintegration process will operate more efficiently for high-frequency words than for low-frequency words because high-frequency words have representations in long-term memory that are more easily accessed by partial information. Thus, frequency will be relatively unimportant for items early in the list (because they are read directly out of short-term memory), but will affect the recall of items later in a list (because at retrieval they have been subject to significant degradation within the short-term store and thus undergo redintegration). The model predicts that the difference in performance between high-frequency words and low-frequency words will increase systematically with serial position. We assessed this prediction here by looking at serial position functions in the immediate recall of supraspan lists.

Method

Participants. The participants were 30 first-year psychology students (22 women and 8 men) at the University of York, who participated in compliance with a course requirement.

Materials. Two sets of eight words were used in the experiment: the short, high-frequency words and the short, low-frequency words used in Experiment 1.

Procedure. The participants were tested individually in a single session lasting approximately 40 min. In this experiment a trial consisted of the presentation of a list of seven words at a rate of 1 item per second. The apparatus and software used to present the stimuli were as described for Experiment 1. Participants were instructed to attempt to recall the items in their correct order as soon as the final item had been presented. They were asked to say "blank" for any items they could not remember.

At the beginning of the session, each participant was given three practice trials with a set of words not used in the experiment. After this, the experimental conditions were tested, with order of testing counterbalanced across participants.

In each condition, the participants initially listened to and repeated each of the eight words for that condition to check their audibility. They were then presented with 25 lists of seven words. Each list was constructed by drawing at random, without replacement from the pool of words for that condition. After all the lists for one condition had been presented, the participant's speech rate was measured by having the participant repeat each word in the set for that condition 10 times, as quickly as possible, until told to stop. The participant's voice was recorded during the speech rate task, and the time taken to say each word 10 times was measured from the visual wave form after the audio recording had been transferred to a computer. The mean of these times was converted into a speech rate in items per second for that condition.

Results

The mean speech rate for the high-frequency words was 3.144 words per second ($SD = 0.618$), whereas for the low-frequency words it was 3.109 words per second ($SD = 0.689$). This difference was not significant, $t(29) = 0.652, p > .1$.

Each trial was scored for the number of words correctly recalled in serial position. These data, collapsing across participants and lists, are presented in Figure 3.

These data were subjected to a split-plot ANOVA in which the repeated measures variables were word frequency and serial position, and the single between-groups variable, with two levels, was order of test. The main effect of order was

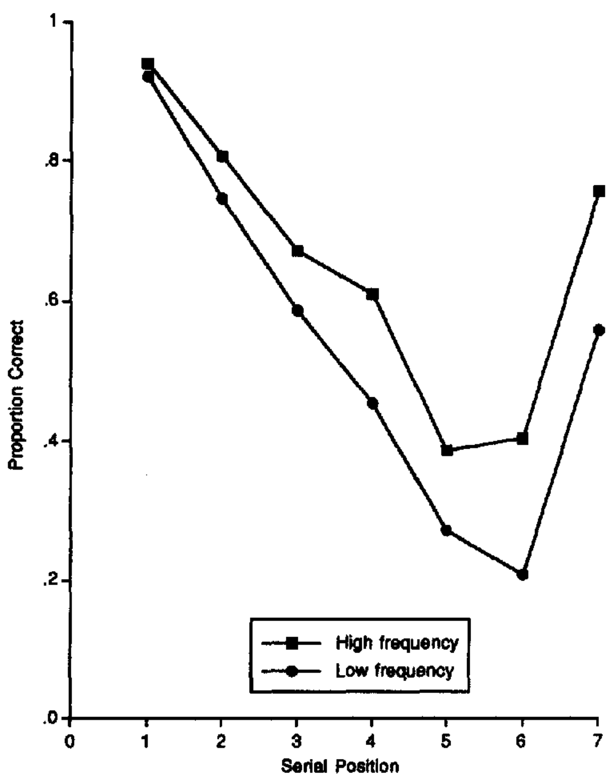

Figure 3. Mean proportion correct recall as a function of serial position for high- and low-frequency words in Experiment 2.

not significant, $F(1, 28) = 0.00$, $MSE = 0.19$, *ns*, and it did not interact with any other variable. There was a significant effect of word frequency, $F(1, 28) = 39.64$, $MSE = 0.04$, $p < .001$; of serial position, $F(6, 168) = 142.66$, $MSE = 0.02$, $p < .001$; and a significant interaction between these two variables, $F(6, 168) = 7.96$, $MSE = 0.01$, $p < .001$. As can be seen from Figure 3, the form of this interaction, with an increasing effect of word frequency from early to late positions, takes the form predicted by the multinomial processing tree model. An evaluation of the goodness of fit provided by the model for the data obtained here is included in Appendix B.

The multinomial processing tree model also makes predictions about the types of errors that should occur in recall. If, as we have argued, participants find it more difficult to redintegrate the degraded memory traces of low-frequency words than of high-frequency words, then participants should more often omit items or recall erroneous items in the low-frequency lists. The model has no explicit mechanism for predicting order errors (occasions when an item from the list is recalled in the wrong position), though with additional assumptions the model could be extended to deal with such errors. However, without additional assumptions, the model does not predict that there should be any difference in the number of order errors made on high- and low-frequency lists. The data collected in this experiment were used to assess these predictions.

Each recall error was classified into three mutually exclusive categories: order errors (recalling an item from the list in an incorrect position), omission errors (saying "blank" when an item could not be recalled), and errors of commission (recalling an item that was not actually presented in the list). The error rates for each of these categories reported are the mean number per trial, averaged over participants.

The number of order errors did not differ between the high-frequency lists (1.211) and the low-frequency lists (1.271), $t(29) = 0.648$, *ns*, two-tailed paired. However, as predicted, errors of omission were much more frequent for low-frequency lists (1.863) than for high-frequency lists (1.168), $t(29) = 6.171$, $p < .01$. These results support our prediction that unsuccessful redintegration of the low-frequency word traces increases the likelihood of the participants simply saying "blank" when they cannot recall an item. There were also slightly more errors of commission made on the low-frequency lists (0.137) than on the high-frequency lists (0.083), $t(29) = 2.209$, $p < .05$. This latter result is also consistent with redintegration being less likely to succeed for the low-frequency than for the high-frequency word lists.

These errors of commission were also broken down into intraset intrusions (intrusions of the eighth set item that was not actually presented on the trial) and extraset intrusions (intrusions of an item that was not in either of the experimental sets of eight words). Because the number of commission errors was small, the following figures represent the mean number of items per participant, collapsing across all 25 lists in each condition. Intrusions between word sets (high- and low-frequency sets) did not occur. For low-frequency lists, there were roughly equal numbers of intraset (1.97) and extraset intrusions (1.47). However, for high-frequency lists, there were 2.07 intraset intrusions and no extraset intrusions. In other words, for the high-frequency lists, if a participant ever said an item that was not in the list, they inserted the word from the high-frequency word set that had not actually been presented on that trial. It is also striking that a vast majority of the extraset intrusions reported in the low-frequency word trials were substitutions of similar-sounding high-frequency words. The three most common errors of this type (*foul* for *foal, truth* for *truce,* and *list* for *lisp*) accounted for 72.5% of the extralist intrusions on the low-frequency lists.

Our findings concerning order errors contrast with some recent findings of DeLosh and McDaniel (1996). DeLosh and McDaniel compared participants' ability to perform free recall of high-frequency and low-frequency word lists after a filled delay. They also required participants to order items (when all items were represented) to correspond to the order of presentation. Participants were more successful in ordering high-frequency words than low-frequency words, and it was argued that order information was more efficiently encoded for high-frequency words. The contrast between our results and those of DeLosh and McDaniel probably depends on a number of procedural differences between experiments. Our claims for the importance of a redintegration process that operates on degraded phonological memory traces are specific to immediate serial recall: We would not expect equivalent results in a delayed memory-ordering response, in which item errors are eliminated because all participants have to do is reorder the items that are presented and reliance on a phonological code is likely to be minimal because of the filled delay between presentation and response.

Discussion

The results of this experiment replicate those of Experiment 1 in clearly showing an effect of frequency on short-term serial recall that is independent of speech rate. Furthermore, they indicate that the effect of frequency increases systematically across the nonrecency serial positions in a list.

The suggestion by Watkins (1977) that frequency affects short-term recall because items early in the list are recalled from long-term memory is challenged by the results of this experiment. This theory predicts that frequency should have a greater effect on serial recall in early positions in the list (because early items are recalled from long-term memory, and high-frequency items are coded more efficiently by long-term memory). Contrary to this prediction, the present experiment has demonstrated the opposite pattern, with an increasing effect of frequency from early to later serial positions. The persistence of a difference between the high-frequency and low-frequency words in the late serial positions is particularly damaging for the model advocated by Watkins.

We argued that Schweickert's (1993) model predicts an increasing effect of frequency across serial position, as has been observed in this experiment, for two reasons. First, because low-frequency words might take longer to say, items in the later serial positions would degrade more during the output of the list when the items were of low frequency than when they were of high frequency. However, it seems likely that this effect would always be small, and indeed in this experiment the difference in speech rate between the two conditions was extremely small and far from significant. We conclude that the difference in speech rate between the high- and low-frequency words cannot account for the differences in patterns of recall observed.

The second reason that Schweickert's (1993) model predicts an interaction between frequency and serial position is through an effect of frequency on the redintegration process used when traces from the short-term phonological stores have decayed beyond immediate recognition. As suggested earlier, the effect of frequency on the redintegration process is assumed to arise from the greater accessibility of long-term memory representations of high-frequency words than of low-frequency words.

Experiment 3

Experiment 2 showed, as predicted by the multinomial processing tree model, that word frequency has an effect on recall that increases across the nonrecency serial positions. We also showed (see Appendix B) that the model provided a good quantitative fit to the data.

One possible objection to these findings, however, is that we may simply have a ceiling effect operating. It might be argued that the nearly perfect performance at Serial Position 1 in Experiment 2 makes it appear artificially that the effects of word frequency interact with serial position. A direct test of this can be undertaken by comparing the effects of word frequency across serial positions for words of different lengths, because longer words should lead to lower levels of recall. We wished to see if the same pattern (with increasing effects of word frequency across the nonrecency portions of the serial position curve) would be obtained for sets of words of different lengths with different levels of absolute performance.

In Experiment 3, we therefore compared recall of short- and medium-length words of high and low frequency and hoped to replicate the interaction between frequency and serial position in both sets of words. We also predicted an interaction between word length and serial position, with a steeper decrease across serial positions for the longer words. This follows from the model because longer words early in a list should lead to greater degradation of the traces of later words, and so parameter I (the probability of direct recall) should decrease more steeply across serial positions for longer words. As noted above, just such a pattern was found by Cowan et al. (1992).

Method

Participants. The participants were 20 undergraduate students (14 women and 6 men) at the University of York.

Materials. Four sets of eight words were used as stimuli in this experiment. These were the high- and low-frequency words of short and medium spoken duration used in Experiment 1.

Procedure. The participants were tested individually in a single session, lasting approximately 60 min. Each participant took part in four conditions with 25 trials in each condition: short high-frequency and medium high-frequency words; short low-frequency and medium low-frequency words. The apparatus and software used to present the stimuli were identical to those used in Experiments 1 and 2. Each trial consisted of the presentation of a list of seven words at a rate of one item every second. The order of testing for both variables, frequency and length, was counterbalanced across participants.

Initially for each condition, the participants listened to the eight words for that condition, one at a time, and were asked to repeat them back to check their audibility. They were then presented with 25 lists of seven words, each list having been constructed by drawing at random without replacement from the pool of eight words for that condition. For each list, after the final item had been presented, participants were asked to recall as many items as possible in their correct order and to say "blank" to any that they could not remember. At the end of each condition, the participant's speech rate for that condition was measured. They were asked to repeat each word in the set 10 times, as quickly as possible, and the time taken to do this was recorded by using a stopwatch and converted into a speech rate in words per second.

Results

The mean speech rates, in words per second, for the four sets of words were as follows: short high-frequency, 3.658 (SD = 0.467); short low-frequency, 3.615 (SD = 0.458); medium high-frequency, 2.638 (SD = 0.250); and medium low-frequency, 2.429 (SD = 0.209).

Each trial was scored for the number of words recalled in correct serial position. These data, collapsing across participants and lists, are presented in Figure 4.

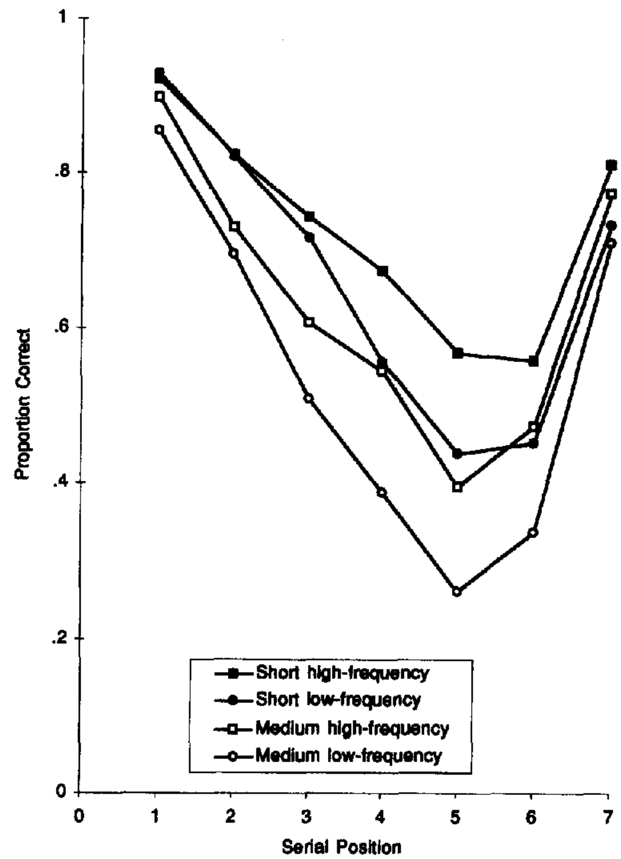

Figure 4. Mean proportion correct recall as a function of serial position for short- and medium-length words of high and low frequency in Experiment 3.

The recall scores were subjected to a four-way split-plot ANOVA, in which the within-subject variables were word length, word frequency, and serial position, and the between-subjects variable was order of test, with four levels. The main effect of order was not significant, $F(3, 16) = 0.14$, $MSE = 278.73$, *ns*, and it did not interact with any other variables. There were significant main effects of word frequency, $F(1, 16) = 21.33$, $MSE = 26.07$, $p < .001$; of word length, $F(1, 16) = 59.00$, $MSE = 18.60$, $p < .001$; and of serial position, $F(6, 96) = 39.16$, $MSE = 40.95$, $p < .001$. There were also, as predicted by the model, significant interactions between frequency and serial position, $F(6, 96) = 6.47$, $MSE = 5.03$, $p < .001$, and of word length and serial position, $F(6, 96) = 3.70$, $MSE = 3.70$, $p < .01$. No other effects in this analysis were significant.

A three-way split plot ANOVA was also carried out on the speech rate data, in which the within-subject variables were word frequency and word length, and the between-subjects variable was order of test, with four levels. The main effect of order was not significant, $F(3, 16) = 0.93$, $MSE = 0.32$, *ns*, and it did not interact with any other variables. This analysis revealed significant main effects of word frequency, $F(1, 16) = 11.26$, $MSE = 0.03$, $p < .01$, and of word length, $F(1, 16) = 261.97$, $MSE = 0.09$, $p < .001$. No other effects in this analysis were significant. Because significant effects were obtained for both word frequency and word length in the speech rate data, this raises the possibility that effects of word frequency and word length in the initial analysis of the serial recall data may be a result of speech rate differences.

To determine whether there are effects of word frequency and word length on memory, over and above the effects that are due to speech rate differences between the conditions, we subjected the serial-recall scores to an ANCOVA, with speech rate as the covariate; other aspects of the analysis were the same as before. This analysis allowed us to control for the linear effects of differences in speech rate on serial-recall scores. If differences in serial recall for high-frequency and low-frequency words or for short and medium words remained after the effects of speech rate have been controlled, this would provide good evidence that the remaining significant effects are not attributable to differences in speech rate.

The ANCOVA revealed that there was a significant effect of word frequency, $F(1, 15) = 19.63$, $MSE = 25.07$, $p < .001$, confirming that frequency has effects on serial recall that remain when the linear effect of speech rate is removed. It also revealed a main effect of word length, which although weaker than in the initial ANOVA, $F(1, 15) = 6.78$, $MSE = 19.02$, $p < .05$, is still significant. This suggests that even though the majority of the word-length effect is attributable to differences in speech rate, a part of the effect reflects differences between the short- and medium-length words that are not attributable to linear effects of differences in speech rate. The interactions of word frequency with serial position, word length with serial position, and word frequency and word length with serial position, were the same as those obtained in the ANOVA, because the values of the covariate do not differ across serial position.

For Serial Position 1, there was little or no effect of word frequency on recall for short words, consistent with what was found in the previous experiment. However, for medium words there was a significant effect of word frequency at Serial Position 1 (test of two proportions, $z = 2.01, p < .05$). Hence, the significant interaction between word frequency and serial position was not due to the lack of an effect at Serial Position 1 and the presence of an effect elsewhere.

As for Experiment 2, a detailed evaluation of the goodness of fit provided by the multinomial tree model for the data obtained here is included in Appendix B.

Discussion

The results of Experiment 3 provide a useful replication of those of Experiment 2. The form of the interaction between serial position and frequency for both the short- and medium-length words is nearly identical in this experiment. This pattern shows that the effect we observed in Experiment 2, which used short words only, is not simply an artifact of a ceiling effect at Position 1.

The observed levels of performance observed in this experiment fit extremely well with the predictions derived from the multinomial processing tree model (see Appendix B). These data provide strong support for the notion of a redintegration process that operates on degraded traces retrieved from the short-term store once an initial attempt to identify the item has failed.

Experiment 4

The multinomial processing tree model received considerable support from Experiments 2 and 3. The prediction from the model, of an increasing effect of word frequency from early to later serial positions, was confirmed. The model also provides a close quantitative fit to the results obtained. However, it might be argued that the qualitative pattern of results obtained can be explained more simply in terms of the operation of a range effect: As performance levels decline, frequency has a greater effect on performance.

To test the range effect explanation, we decided to look for a manipulation that would depress performance. The manipulation we chose was backward recall. It is well recognized that backward serial recall is generally considerably harder than forward recall (e.g., Cowan et al., 1992). The mechanisms underlying backward recall are not well understood, though they appear to differ in important ways from the mechanisms of forward recall. Li and Lewandowsky (1993) found that backward recall was unaffected by verbal intralist distracters, whereas forward recall was badly harmed. Conversely, a spatial distractor task disrupted backward but not forward recall (Li & Lewandowsky, 1995). From these, and other results, Li and Lewandowsky have argued that backward recall depends on different retrieval processes from those of forward recall, with backward recall depending heavily on a visual–spatial retrieval strategy, whereas forward recall depends on an auditory–verbal retrieval strategy. Farrand and Jones (1996) found that backward recall was inferior to forward recall only when participants were required to recall both item and order information. When memory items were represented at recall, so that participants merely had to select items either in their order of occurrence or in the reverse order, the difference between backward and forward recall was eliminated. Farrand and Jones also argued that backward recall depends on different retrieval mechanisms from those of forward recall, though they dispute Li and Lewandowsky's suggestion that backward recall is critically dependent on a visual–spatial retrieval strategy.

These considerations led to an interesting prediction. We have argued that word frequency has effects on short-term memory performance because it affects a redintegration process that operates (on traces held in a phonological code) at retrieval. If backward recall depends on different retrieval mechanisms from those of forward recall, we might expect word frequency to have less of an effect on backward than on forward recall. However, we can confidently expect that backward recall will show a steeper decline in performance during the recall of successive items than in the case of forward recall. On the basis of a simple range effect explanation, we would therefore expect frequency to have as great, or possibly a greater, effect on performance in backward recall than in forward recall. Experiment 4 tests these opposing predictions.

Method

Details of the method were closely modeled on those for Experiment 2. Once again, participants listened to and recalled seven-item lists of either short high-frequency words or short low-frequency words. The words used and their mode of presentation were identical to those used in Experiment 2. The one difference was that for half of the trials participants were required to recall the lists in reverse order, and for the other half of the trials, participants were required to use forward recall. The experiment was divided into four blocks of 25 trials. Half of the participants performed forward recall for the first two trial blocks and backward recall for the second two blocks. This order was reversed for the remaining half of the participants. Each participant received alternating blocks of high- and low-frequency trials. Order of alternation was balanced both within and between manipulations of recall order.

Participants. A total of 28 participants (24 women and 4 men) took part in the experiment. All were first-year psychology undergraduates at the University of York who participated in return for course credit.

Procedure. As in the earlier experiments, before testing commenced, participants were asked to listen to and to repeat each of the words that would appear in the experiment. They were instructed that as soon as they heard the last word in the list, they should recall the stimuli either in the same order in which they were presented or in the reverse order, depending on the specific instructions given before each trial block. Participants responded orally, and we recorded each response as correct or incorrect. As before, participants were instructed to say "blank" if they could not recall an item from a particular serial position. On completion of the four experimental trial blocks, an estimate of each participant's speech rate was made for both the high- and low-frequency words by using the same procedure as in Experiment 3.

Results

Although there appeared to be very little difference in the average speech rate for low-frequency (2.66 words per second, $SD = 0.62$) and for high-frequency words (2.85 words per second, $SD = 0.60$), a one-tailed paired t test showed that the difference was reliable, $t(27) = 2.27$, $p < .05$.

For the recall task, only words recalled in the correct serial position were scored as correct. The results, collapsing across participants and lists, are shown in Figure 5. It is clear from Figure 5 that the pattern obtained in Experiments 2 and 3 for forward recall, with an increasing effect of word frequency from early to late serial positions, is replicated here. However, it is also clear that the pattern for backward recall is quite different. For backward recall, there is very little difference between the high- and low-frequency words and no sign of any increase in the size of the frequency effect across serial positions.

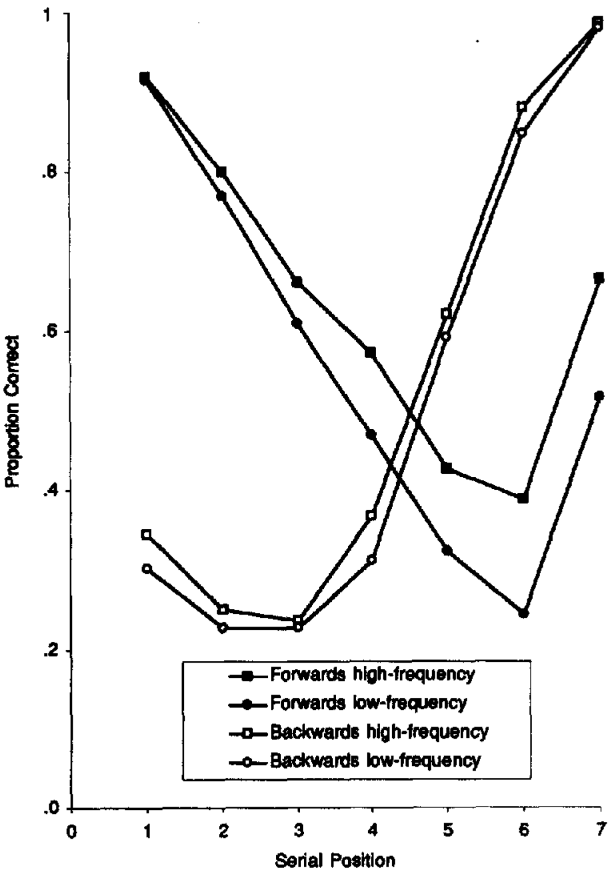

Figure 5. Mean proportion correct recall as a function of serial position and direction of recall for high- and low-frequency words in Experiment 4.

The recall data were entered into an ANOVA with three within-subject variables, recall direction (forward vs. backward), word frequency (high vs. low), and serial position at presentation (1–7); and one between-subjects variable, test order (1–4). The analysis revealed a main effect of word frequency, $F(1, 24) = 18.89$, $MSE = 20.13$, $p < .01$, indicating that overall high-frequency words were better recalled. There was also a significant main effect of direction of recall, $F(1, 24) = 13.54$, $MSE = 56.44$, $p < .01$, confirming that as expected, backward recall was harder than forward recall. The main effect of serial position was also significant, $F(6, 144) = 65.65$, $MSE = 16.95$, $p < .01$, though little interpretation of this effect can be offered because it results from an amalgamation of the forward and backward recall curves. The main effect of test order was not significant, $F(3, 24) = 0.07$, $MSE = 212.43$, *ns*, nor did it interact significantly with any other variables.

The interaction between recall direction and serial position, $F(6, 144) = 311.19$, $MSE = 12.28$, $p < .001$, was highly significant, reflecting the cross-over pattern between the forward and backward serial position curves. The Recall Direction × Word Frequency interaction, $F(1, 24) = 9.70$, $MSE = 9.59$, $p < .01$, was also significant, reflecting the fact that the frequency effect was greater in forward than in backward recall. There was also a significant two-way interaction between word frequency and serial position, $F(6, 144) = 2.63$, $MSE = 5.14$, $p < .01$, but interpretation of this interaction is not possible given the significant three-way interaction between recall direction, word frequency, and serial position, $F(6, 144) = 3.42$, $MSE = 4.69$, $p < .01$. This latter interaction clearly confirms what can be seen from Figure 5: that there is an increasing effect of word frequency over serial position in forward but not in backward recall.

This three-way interaction was further explored by examining forward and backward recall performance separately. Because test order had not entered into any of the significant effects in the previous analysis, this variable was dropped from subsequent analyses. For forward recall, both the main effects were significant for serial position, $F(6, 162) = 105.03$, $MSE = 15.15$, $p < .01$, and for frequency, $F(1, 27) = 24.44$, $MSE = 17.37$, $p < .01$, and most critically, there was also a significant Word Frequency × Serial Position interaction, $F(6, 162) = 4.67$, $MSE = 5.79$, $p < .01$. In contrast, for backward recall, the effect of word frequency was not significant, $F(1, 27) = 3.52$, $MSE = 13.79$, $p = .07$, nor was there any sign of an interaction between word frequency and serial position, $F(6, 162) = 0.57$, $MSE = 4.45$, *ns*. The main effect of serial position was highly significant, $F(6, 162) = 217.29$, $MSE = 15.39$, $p < .01$, reflecting the large differences in performance across serial positions observed in the backward recall condition.

Because speech rate for the high-frequency words was slightly, though significantly, higher than that for the low-frequency words, it was of interest to assess the extent to which speech rate may have contributed to the frequency effect found in the recall data. An ANCOVA on all of the data, with speech rate as the covariate, showed that although the effect of frequency was weaker, it remained significant, $F(1, 23) = 11.09$, $MSE = 17.64$, $p < .01$. Thus, as in the earlier experiments in this series, differences in recall between high- and low-frequency words were not attributable to the linear effects of differences in speech rate between these sets of items. Separate ANCOVAs, with speech rate as the covariate, were also conducted on the forward and backward recall data. For forward recall, the size of the frequency effect was only slightly weaker, $F(1, 26) = 16.04$, $MSE = 16.19$, $p < .01$. However, the marginal main effect of frequency for backward recall was far from significant when speech rate was included as a covariate, $F(1, 26) = 0.87$, $MSE = 11.91$, *ns*.

The present experiment, like the previous experiments in this series, involved repeated presentations of items drawn from the same small pools of words (8 words per condition). It might be argued that with such a procedure the effect of word frequency should become attenuated over trials (because with repeated presentations participants become familiar with both the high- and the low-frequency words used in the experiment). However, if, as we have argued, the effects of word frequency depend on differences in the lexical representation of the phonological forms of the words used, it would be expected that even repeated presentations would be insufficient to change the nature of such representations within a single experimental session. To assess this, we decided to examine the magnitude of the effect of word frequency across trials in the experiment. For the forward recall procedure, the number of words correctly recalled for each high-frequency list and each low-frequency list was calculated. These scores were then entered into an ANOVA in which the variables were word frequency (with two levels) and trial number (1–25). This analysis revealed a main effect of word frequency, $F(1, 27) = 24.44$, $MSE = 4.86$, $p < .001$, but no effect of trial number, $F(24, 648) = 0.842$, $MSE = 2.70$, *ns*, and no interaction between word frequency and trial number, $F(24, 648) = 0.65$, $MSE = 2.65$, *ns*. Thus, there is no change in performance levels across trials in this experiment, and the effect of word frequency neither decreases nor increases as participants become more familiar with the words used. This pattern is in line with what we would expect given the view that word frequency has effects because of differences in the lexical representation of high- and low-frequency words.

For forward recall in this experiment, as in Experiments 2 and 3, word frequency had a clear effect on performance of the form predicted by the multinomial processing tree model. Accordingly, for the forward recall data, a detailed evaluation of the goodness of fit provided by the model is included in Appendix B. As for the earlier experiments, the model fits the data well, with the possible exception of the last serial position (see Table B5 in Appendix B).

Discussion

The results of the present experiment show that a range effect explanation for the interaction between word frequency and serial position in Experiments 2 and 3 is not tenable. In backward recall, there are large declines in recall accuracy in the medial serial positions, but at no point is there a reliable effect of word frequency. Conversely, for forward recall, where performance is at a higher level, word frequency has an effect that increases systematically from early to later serial positions.

The complete absence of a frequency effect for backward recall is perhaps somewhat surprising. This finding lends support to the view that backward recall depends on different retrieval mechanisms from those for forward recall. However, the present experiment does not really clarify the nature of the retrieval mechanisms involved in backward recall. Li and Lewandowsky (1993, 1995) have argued that in backward recall participants depend on a visual–spatial retrieval strategy. In their own experiments, Li and Lewandowsky used visual presentation in which the use of a visual–spatial strategy may seem plausible. However, in the present experiment, with auditory presentation and spoken recall, we doubt that a visual strategy is likely to have been used. One possibility, if participants are placing less reliance on a phonological strategy in backward recall, is that a semantic retrieval strategy is used. It is clear that further experimental studies are needed to clarify the differences in retrieval processes between forward and backward recall.

Finally, the absence of a word-frequency effect for backward recall provides further evidence that the effect of word frequency on serial-recall performance is independent of any effects that word frequency may have on speech rate (cf. Experiments 1–3). Cowan et al. (1992) found roughly equivalent effects of word length for both forward and backward recall, which they attributed to the effects of word length on the rate of output during recall. The absence of an effect of word frequency on backward recall, when such an effect is present in forward recall, therefore provides further evidence that the effect of word frequency on serial recall is not attributable to differences in speech rate.

General Discussion

The results from Experiments 1–4 show that word frequency has effects on short-term serial recall that are independent of speech rate. In addition, we have also shown that the effect of word frequency on recall increases with serial position in forward recall, at least until the last positions.

The fact that serial position and frequency interact in this manner goes against the claim that participants recall items early in the list from long-term memory and later items from short-term memory (Watkins, 1977). If this was the case, assuming frequency does not affect retrieval from the short-term store, we would expect to see a decreasing effect of frequency across serial positions. In contrast, a two-component model, in which frequency influences the secondary process of redintegration, receives strong support. We have argued that word frequency influences the redintegration of partially decayed traces retrieved from a short-term store (Brown & Hulme, 1995; Hulme et al., 1991; Schweickert, 1993). In this view the phonological representations of high-frequency words are more accessible or better specified and hence support the process of redintegration more effectively than the representations of low-frequency words.

We must consider the nature of the redintegrative process and the representations on which it operates. It has often been argued that short-term memory involves processes common to speech perception and speech production (e.g., Ellis, 1980; Schweickert, 1993). This suggests that the representations underlying the redintegration process are the same, or at least similar to, those that exist primarily to subserve the perception and production of speech.

Schweickert (1993) argued for two distinct redintegrative processes that are analogous to processes used to repair errors in speech production. The two processes operate at different levels in parallel, one acting to transform the degraded representation into a word and the other acting to transform such a representation of a word into a string of phonemes. This model assumes that items are maintained in a short-term store in some form of phonological or articulatory code. These representations are then used to feed the speech production processes that give rise to the frequency effect we have observed and modeled in these experiments.

These ideas are close to the proposal made by Hulme et al. (1991) that verbal short-term memory processes might usefully be thought of as a by-product of processes involved in speech perception and production. Hulme et al. went on to relate this idea to a connectionist model of short-term memory and verbal rehearsal that interfaced separate speech perception and speech production networks (Brown, 1990). Within this framework, as within Schweickert's (1993), the redintegrative process is an integral part of the normal processes that operate during speech production. More specifically, redintegration is the clean up of noisy representations within a network that is characteristic of all connectionist models.

This general approach can give an admittedly speculative account of the pattern of results obtained in our experiments. Such a model would predict an interaction between frequency and serial position if the cleanup process in the speech production net plays a more important role as traces decay. Decay within the network can be thought of as the disintegration of the pattern of activation representing that item. The clean up processes attempt to reconstruct the whole pattern on the basis of some fragment or noisy version of the pattern. The more noisy the pattern is (i.e., the more decayed it is) the less likely it is that the clean up processes will correctly reconstruct the pattern. At the same time, the clean up process would be expected to be more effective for high-frequency words whose representations are better specified than the representations of low-frequency items.

In summary, in this article we have demonstrated an effect of word frequency on immediate serial recall and have applied a multinomial processing tree model that accounts well for the data. This model assumes that degraded short-term memory traces are reconstructed in a manner similar to the repair of speech errors in speech production systems. We have also suggested that it may be fruitful to consider the extent to which immediate serial recall of auditorily presented lists depends on the operation of speech perception and speech production mechanisms.

References

Baddeley, A. (1986). *Working memory.* London: Oxford University Press (Clarendon).

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36*(A), 233–252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology, 39,* 129–149.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing tree models of source monitoring. *Psychological Review, 97,* 548–564.

Brown, G. D. A. (1990). A neural-net model of human short-term memory. In L. B. Almeida & C. J. Wellekens (Eds.), *Neural networks: Lecture notes in computer science* (Vol. 412, pp. 175–185). Heidelberg, Germany: Springer-Verlag.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594–621.

Chechile, R. (1987). Trace susceptibility theory. *Journal of Experimental Psychology: General, 110,* 203–222.

Chechile, R., & Meyer, D. L. (1976). A Baysean procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology, 13,* 269–295.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31,* 668–684.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31,* 1–17.

Cox, R., Hulme, C., & Brown, G. D. A. (1992). STM Experimenter: Using Hypercard and MacRecorder in short-term memory experiments. *Behavior Research Methods, Instruments and Computers, 24,* 575–579.

Craik, F. I. M. (1968). Two components in free recall. *Journal of Verbal Learning and Verbal Behavior, 7,* 996–1004.

Crowder, R. G. (1978). Memory for phonologically uniform lists. *Journal of Verbal Learning and Verbal Behavior, 17,* 73–89.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word–frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1136–1146.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19,* 624–634.

Farrand, P., & Jones, D. M. (1996). Direction of report in spatial and verbal serial short-term memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 140–158.

Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80,* 363–374.

Howes, D. H. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29,* 296–305.

Hu, X. (1991). General program for processing tree models [Computer program]. Irvine, CA: University of California, School of Behavioral Sciences.

Hu, X., & Batchelder, W. H. (1992). *The statistical analysis of general processing tree models with the EM algorithm* (Tech. Rep. No. MBS 92-06). Irvine, CA: University of California, School of Behavioral Sciences.

Hulme, C., & Mackenzie, S. (1992). *Working memory and severe learning difficulties.* Hove, UK: Erlbaum.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

Hulme, C., Roodenrys, S., Brown, G. D. A., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86,* 527–536.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levin, J. R. (1985). Some methodological and statistical 'bugs' in research in children's learning. In M. Pressley and C. J. Brainerd (Eds.), *Cognitive learning and memory in children* (pp. 205–234). New York: Springer-Verlag.

Li, S.-C., & Lewandowsky, S. (1993). Intralist distractors and recall direction: Constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 895–908.

Li, S.-C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 837–847.

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word frequency paradox in recognition. *Memory & Cognition, 10,* 33–42.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398–422.

Quinlan, P. T. (1992). Oxford psycholinguistic database [Computer software]. Oxford, England: Oxford University Press.

Reed, T. R. C., & Cressie, N. A. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

Roodenrys, S., Hulme, C., Alban, J., Ellis, A., & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22,* 695–701.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 167–175.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 419–425.

Sperling, G., & Speelman, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 152–202). New York: Academic Press.

Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society, 26,* 293–296.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534.

Wright, C. C. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition, 7,* 411–419.

(Appendixes follow)

Appendix A

Stimulus Sets Used in Experiment 1

High-frequency words			Low-frequency words		
Short	Medium	Long	Short	Medium	Long
hour	area	administration	foal	brigand	preposition
colour	statement	professional	vow	curfew	complication
game	treatment	opportunity	truce	decoy	rigidity
fear	religion	population	vet	deluge	aviator
view	research	society	crock	bequest	conspirator
art	justice	situation	elf	hybrid	contamination
unit	country	university	dolt	apology	technicality
order	council	individual	lisp	acidity	causality

Appendix B

Evaluating the Goodness of Fit of the Multinomial Processing Tree Model

In the multinomial processing tree model for immediate recall described in Schweickert (1993), the mental processes carried out during the recall of a single item are represented as branches in a tree (Figure 2). With probability I, an intact representation of the item exists, and the item is recalled directly and correctly. Otherwise, the representation is degraded and is correctly reconstructed with probability R. With probability $1 - R$, the redintegration process fails, and an error occurs in recalling the decayed trace. The probability of correct redintegration depends on the person's knowledge of the language and of the particular items used in the experiment. A detailed description of redintegration is given in Sperling and Speelman (1970).

To apply the processing tree model here, we assume that the probability that the representation of an item is intact depends on the serial position of the item but not on its frequency. Likewise, we assume that the probability of correct redintegration depends on the frequency of the item but not on its serial position. Given these assumptions, then the probability of correct recall of an item in serial position i with word-frequency level j (where $j = 1$ for high-frequency words, and $j = 2$ for low-frequency words) is

$$P_{ij} = I_i + (1 - I_i) R_j \tag{B1}$$

The model makes a precise prediction about the form of the interaction between serial position and word frequency, and a better test than the omnibus F from the ANOVA can be made by constructing a contrast for the specific prediction (Levin, 1985, p. 217). The model predicts that the greater the effect of serial position, the greater will be the effect of frequency. In principle, one test of this interaction can be made for every pair of serial positions, but some of these tests would be redundant. The contrast described below has been constructed by combining all such tests; it turns out to have the same form as the test of a linear trend. This contrast tests directly the prediction from the model that the difference between high- and low-frequency words will increase systematically across serial positions.

Contrast Testing the Multinomial Tree Model

Let D_i be the difference in the probability of correct recall for the high- and low-frequency words at position i. From Equation B1, the model predicts that this difference will increase as I_i increases. Typically, recall is higher at the beginning and at the end of the list. Then the prediction is for the difference to increase toward the middle of the list and then to decrease. Suppose for convenience that the positions are renumbered so the position with highest recall is 1, the second highest is 2, and so on. With the new numbering, if D_1 is used as a reference, $D_1 < D_2$, $D_1 < D_3$, and so on. A contrast based on D_1 as the reference for the n positions would be

$$(D_1 - D_2) + (D_1 - D_3) + \ldots + (D_1 - \mathrm{D}_n).$$

The model predicts this to be negative. Of course, there is nothing special about using D_1 as a reference. The analogous contrast based on D_2 as a reference would be

$$-(D_2 - D_1) + (D_2 - D_3) + \ldots + (D_2 - D_n).$$

The difference in the first expression is negative because the model predicts $D_2 > D_1$.

We can continue in this way forming one contrast by using each position as a reference. If we add all n of these contrasts, the result is a combined contrast that uses each position as a reference exactly once. Dividing every term by 2 simplifies the result to

$$(n - 1)D_1 + (n - 3)D_2 + (n - 2i + 1)D_1 + \ldots - (n - 1)D_n. \tag{B2}$$

The model using Equation 1 for every serial position predicts this combined contrast to be negative.

To calculate the value of the contrast, the serial positions must be rank ordered by the probability of correct recall. The model predicts a rank order for the serial positions. The position with the largest value of the parameter I is first, the position with the second highest value of I is second, and so on. The value of the contrast for the model was calculated by using the rank ordering of position determined in this way.

Table B1
Parameter Estimates for Experiment 2

Parameter	Value
I_1	.906
I_2	.702
I_3	.505
I_4	.371
I_5	.117
I_6	.073
I_7	.539
R_1	.355
R_2	.149

Note. Goodness-of-fit values are $G^2 = 28.770$, $df = 5$, and $r^2 = .988$, $df = 5$. I_1–I_7 indicate estimates of the probability that the item is retrieved directly from short-term memory without redintegration. R_1 is the estimate of the value of the redintegration parameter for high-frequency words. R_2 is the estimate of the value of the redintegration parameter for low-frequency words.

Experiment 2

For the data from Experiment 2, the estimated value of the contrast described above is -19.20. It is negative, as predicted by the model, and significantly different from zero, $t(29) = -3.83$, $p < .001$, two-tailed. The contrast would have the value zero if serial position and word frequency have additive effects on the probability of correct recall. The hypothesis of additivity can be rejected.

Parameters for the model in Equation 1 were estimated with the EM algorithm (Dempster, Laird, & Rubin, 1977; Hu & Batchelder, 1992), using a program by Hu (1991). Parameter estimates are given in Table B1, with two measures of goodness of fit, G^2 and r^2. The parameters were chosen to minimize G^2, a quantity approximately equal to the usual chi-square goodness-of-fit statistic (see, e.g., Reed & Cressie, 1988). It is based on all of the cells, both correct and incorrect responses. The value of r^2 is the squared correlation between the observed and the predicted number of correctly recalled words.

Although the positive value of G^2 indicates a less than perfect fit, it is difficult to know how large a value of G^2 to expect because the chi-square distribution is only a rough guide to the distribution of G^2, due to repeated measurements on the same participants. The large value of r^2 clearly indicates that the model accounts for a large percentage of the variance. Observed and predicted values for the frequency of correct responses are given in Table B2.

Table B2
Observed and Predicted Frequencies of Correct Recall in Experiment 2

Frequency	Serial position 1	2	3	4	5	6	7
High							
Observed	704	606	504	457	291	304	569
Predicted	704	606	511	446	323	302	527
Low							
Observed	690	560	440	340	205	157	419
Predicted	690	560	434	349	187	158	455

Note. Total possible correct score in each cell is 750.

Table B3
Observed and Predicted Frequencies of Correct Recall in Experiment 3

Frequency	Serial position 1	2	3	4	5	6	7
Short words							
High							
Observed	461	412	372	337	284	279	406
Predicted	467	421	380	328	279	280	399
Low							
Observed	465	411	358	278	220	227	367
Predicted	459	402	350	286	224	226	374
Medium words							
High							
Observed	449	365	304	272	198	237	388
Predicted	445	373	305	262	204	237	386
Low							
Observed	428	348	255	194	131	169	355
Predicted	431	341	254	201	128	169	357

Note. Total possible correct score in each cell is 500.

Experiment 3

As for Experiment 2, a specific contrast for testing the model was constructed. The value of the contrast for short words is -32.60. As with Experiment 2, the value is negative, as predicted by the model, and significantly different from zero, $t(19) = -2.91$, $p < .01$, two-tailed. The value of the contrast for long words is -26.90. This value too is negative, and significantly different from zero, $t(19) = -2.59$, $p < .05$, two-tailed.

The observed and predicted values for the model are shown in Table B3. The parameter estimates and the values of G^2 and r^2 are shown in Table B4.

It can be seen from these figures that the model provides an excellent fit to the data. In addition, the parameter values behave very systematically. Parameters R_1 and R_2 are the estimates of the

Table B4
Parameter Estimates for Experiment 3

Parameter	Word length: Short	Medium
I_1	.89	.82
I_2	.73	.57
I_3	.58	.34
I_4	.40	.20
I_5	.23	.00
I_6	.23	.11
I_7	.64	.62
R_1	.43	.41
R_2	.28	.25

Note. The goodness-of-fit values for short words are $G^2 = 8.46$, $df = 5$, and $r^2 = .992$, $df = 5$; for medium words, $G^2 = 3.33$, $df = 5$, and $r^2 = .997$, $df = 5$. I_1–I_7 indicate estimates of the probability that the item is retrieved directly from short-term memory without redintegration. R_1 is the estimate of the value of the redintegration parameter for high-frequency words. R_2 is the estimate of the value of the redintegration parameter for low-frequency words.

Table B5
Observed and Predicted Frequencies of Correct Recall for the Forward Recall Data in Experiment 4

	Serial position						
Frequency	1	2	3	4	5	6	7
High							
Observed	644	560	463	401	299	272	465
Predicted	648	565	471	399	309	271	441
Low							
Observed	641	539	427	328	227	172	362
Predicted	637	534	420	330	220	173	383

Note. Total possible correct score in each cell is 700.

probability of a nonintact high-frequency and a nonintact low-frequency item being correctly reconstructed. These values were estimated separately by the program for the short and medium words, but it is remarkable how closely these estimates agree.

The probability I_i that an item in serial position i is intact decreases linearly with serial position for the nonrecency part of the curve. For Serial Positions 1 to 5, linear regression of I_i against serial position i gives, for short words, $I_i = 1.06 - .16_i (r = -.9992)$. For medium words the equation is $I_i = .99 - .20_i (r = -.9957)$.

The intercept of both regression lines is almost exactly 1; in other words, if the amount of degradation an item has experienced is 0, then the probability that the item is intact is predicted to be 1, as it should be. The probability that an item in the last serial position is intact, I_7, is approximately the same for short and medium words. This is consistent with a finding reported by Baddeley, Lewis, and Vallar (1984; see also Baddeley, 1986, Figure 5.2, p. 83). For written recall of long and short words (in a control condition without articulatory suppression), there was a clear effect of word length on all the serial positions except the last.

The finding here is similar. The observed error frequency for high-frequency words is about the same for short and medium words (94 and 112, respectively). These error frequencies are about the same for low-frequency words (133 and 145, respectively).

Experiment 4

The multinomial tree model was fit to the forward recall data, where there was a significant effect of word frequency. As for Experiments 2 and 3, the specific contrast for testing the multinomial tree model was evaluated. For forward recall, the value of the contrast is −15.36. As for Experiments 2 and 3, this value is negative, as predicted by the model, and significantly different from zero, $t(27) = -3.91$, $p < .001$, two-tailed. For the backward recall condition, the contrast was also evaluated. Its value, averaged over participants, is −0.21, and as we would expect, it is not significant, $t(27) = -.07$, *ns*. This result confirms our conclusion that the effect of word frequency in backward recall, if there is any, does not change with serial position.

The observed and predicted values for the model for the forward recall data are shown in Table 5. The parameter estimates and the values of G^2 and r^2 are in Table 6. As in the earlier experiments, the model fits the data quite well. The stimuli for this experiment are the same as those for Experiment 2, and as would be expected, the parameter values are similar for the two experiments.

Table B6
Parameter Values for the Forward Recall Conditions in Experiment 4

Parameter	Value
I_1	.884
I_2	.696
I_3	.485
I_4	.321
I_5	.119
I_6	.032
I_7	.417
R_1	.366
R_2	.222

Note. Goodness-of-fit values are $G^2 = 8.801$, $df = 5$, and $r^2 = .995$, $df = 5$. I_1–I_7 indicate estimates of the probability that the item is retrieved directly from short-term memory without redintegration. R_1 is the estimate of the redintegration parameter for high-frequency words. R_2 is the estimate of the redintegration parameter for low-frequency words.

Received September 19, 1995
Revision received December 27, 1996
Accepted January 3, 1997 ■