Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis

JEAN SAINT-AUBIN and MARIE POIRIER
University of Laval, Quebec, Canada

In two experiments, the immediate serial recall of lists of words or nonwords was investigated under quiet and articulatory suppression conditions. The results showed better item recall for words but better order recall for nonwords, as measured with proportion of order errors per item recalled. Articulatory suppression hindered the recall of item information for both types of lists and of order information for words. These results are interpreted in light of a retrieval account in which degraded phonological traces must undergo a reconstruction process calling on long-term knowledge of the to-be-remembered items. The minimal long-term representations for nonwords are thought to be responsible for their lower item recall and their better order recall. Under suppression, phonological representations are thought to be minimal, producing trace interpretation problems responsible for the greater number of item and order errors, relative to quiet conditions. The very low performance for nonwords under suppression is attributed to the combination of degraded phonological information and minimal long-term knowledge.

In immediate serial recall, where subjects must recall verbatim a list of items immediately after their presentation, it is well known that long-term knowledge contributes to performance. This contribution is evidenced, among other phenomena, by the better recall of words over nonwords (Besner & Davelaar, 1982; Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995; Richard, 1993). To account for this effect, a number of researchers have put forward what can be called a *reconstruction hypothesis* (Brown & Hulme, 1992, 1995; Hulme et al., 1991; Hulme et al., 1995; Schweickert, 1993). This hypothesis has also been successfully used to handle the effects of most, if not all, other long-term memory factors recently investigated in this context. Hence, it has been applied to word frequency (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994), semantic similarity (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999b), and imageability (Walker & Hulme, 1999).

The reconstruction hypothesis can be described as follows. At the point of recall, phonological representations set up by list presentation are thought to be degraded and

cannot be output directly as responses. Instead, they must undergo a reconstruction process that calls on the long-term representations of the to-be-recalled items. In this process, degraded phonological representations are used as retrieval cues for accessing an acceptable recall candidate. This hypothesis can easily account for the effects of lexicality by assuming that words are better recalled than nonwords because, for words, long-term information is available to support reconstruction (Brown & Hulme, 1992, 1995; Hulme et al., 1991; Hulme et al., 1995; Schweickert, 1993).

Typically, the reconstruction hypothesis is used to interpret performance when a strict serial recall criterion is used. This performance measure factors in item and order information: To be considered correct, an item must be recalled in its exact presentation position. However, this overall measure can present a misleading picture, because some factors influence item and order recall differentially. For instance, word frequency and semantic similarity only influence item recall (Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999b).

Recently, Poirier and Saint-Aubin (1996; Saint-Aubin & Poirier, 1999a, 1999b) introduced two additional assumptions to account for the influence of long-term memory factors on item and order information. More specifically, they suggested that, at recall, (1) phonological representations are output in the appropriate order and (2) order errors are produced by problems in the reconstruction of degraded phonological traces (see, e.g., Schweickert, Guentert, & Hersberger, 1990; Sperling & Speelman, 1970). This version of the proposal—hereafter, called the *retrieval-based hypothesis*—integrates these assumptions with the reconstruction hypothesis in the following manner. List presentation creates a phonological representation of the to-be-recalled items, subject to

degradation. It is assumed that the items are represented in their order of appearance, as, for example, in a memory vector (see Nairne, 1990, for a related idea). Consequently, at recall, phonological representations are output in the appropriate order. The degraded phonological representation of a given item serves as a retrieval cue for accessing the appropriate long-term representation. *Item* recall is influenced by the efficacy of the retrieval process, implying that it is enhanced by factors increasing the accessibility of the appropriate long-term representation. Also, item recall is thought to be hindered by factors increasing degradation of phonological cues. On the other hand, *order* errors are thought to be produced by trace reconstruction problems during the recall process (Lewandowsky & Murdock, 1989; Nairne, 1990; Schweickert et al., 1990; Sperling & Speelman, 1970). For example, if certain phonological features appear in a number of list items, it is easy to see how a given degraded phonological trace could mainly comprise nonspecific features (Schweickert et al., 1990; Sperling & Speelman, 1970). This would entail a higher probability of erroneously interpreting a given trace as one of the other items within the list.

This hypothesis has been successfully applied to the effects of frequency and semantic similarity on item and order information in an immediate serial recall task (Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999b). However, it has not been investigated for lexicality. This test is important for a number of reasons. First, the development of the reconstruction hypothesis is based mainly on lexicality effects. Hence, demonstrating that the retrieval-based hypothesis can handle a more detailed examination of these effects—namely, lexicality effects on item and order information—is an important test of the proposal. Perhaps more important, the retrieval-based hypothesis predicts a different pattern of errors for lexicality than what has been found for frequency and semantic similarity. As will be seen below, this prediction is counterintuitive and constitutes a rather strict test of the proposal. Finally, the predicted pattern of errors would be undetectable with a global performance measure.

Previous results have shown that high-frequency words and semantically similar items are associated with better item recall, while leaving order information unaffected. According to the retrieval-based hypothesis, the item recall advantage occurs because the long-term representations of high-frequency words are easier to access, heightening the probability of correctly interpreting a degraded phonological trace. The same is true of categorized lists where the semantic category shared by the items would enhance the accessibility of the appropriate recall candidate, perhaps by providing an additional retrieval cue. For order information recall, Poirier and Saint-Aubin (1996; Saint-Aubin & Poirier, 1999b) showed that neither frequency nor semantic similarity has an effect, at least not with the order error measures they used. The retrieval-based hypothesis predicts this, because order errors are attributed to reconstruction problems stemming from the loss of distinct phonological features. Because

neither frequency nor semantic similarity influences degradation of phonological traces, order information recall should be left unaffected. More specifically, the supported prediction is that order errors are constant across conditions when expressed as a proportion of recalled items.

However, a strict interpretation of the retrieval-based hypothesis makes a different prediction in the case of lexicality: better order recall for nonwords. Given that no adequate long-term representations are available for nonwords, the reconstruction process, for all practical purposes, is thought not to operate for these items. Because order errors are assumed to be a by-product of reconstruction, the retrieval-based hypothesis predicts that order errors will be extremely infrequent for nonwords. Therefore, the prediction is that there will be a significantly greater proportion of order errors for words than for nonwords. With respect to item information, recall is expected to be better for words than for nonwords. Again, this is because adequate long-term representations are associated with the former but not with the latter. Without adequate prior representations, only a few nonwords would be maintained, perhaps through rehearsal, and these could be output as responses.

EXPERIMENT 1

Here, the effects of lexicality on item and order information were investigated within a standard immediate serial recall paradigm. An articulatory suppression condition was also included as a methodological control and as a further test of the retrieval-based hypothesis. Articulatory suppression is thought to produce greater degradation of phonological traces, either through decay (Baddeley, Lewis, & Vallar, 1984) or through interference (Nairne, 1990). For words, the retrieval-based hypothesis then predicts a greater number of item errors and a greater proportion of order errors; greater degradation of phonological traces implies a less efficient retrieval process and a greater probability of confusing one item for another. In the case of nonwords, because suppression implies that, at recall, fewer phonological traces are intact, a strict interpretation of the retrieval account predicts a greater number of item errors, resulting in an extremely low item recall level. With respect to order errors, the effects of suppression should be minimal, because order errors are attributed to the reconstruction process, which is assumed to be inoperative in the case of nonwords, either with or without suppression. Consequently, similar proportions of order errors should be observed for nonwords under quiet and suppression conditions.

As for methodological considerations, although Richard (1993) showed that the words and nonwords used here were articulated at the same speed by his subjects, it is notoriously difficult to equate words and nonwords in this respect. If pronunciation times are not equivalent, interpretation of the lexicality effect could be difficult, because it is well known that recall is lower for items taking longer to pronounce (Baddeley, Thomson, & Buchanan, 1975; Schweickert & Boruff, 1986). However, it has also

been shown that when a suppression requirement is introduced, recall performance becomes equivalent for items of various length (Baddeley et al.,1984). Consequently, provided lexicality has a sizable effect under suppression, differences in immediate serial recall performance can be attributed to lexicality. Also, after the memory task, the subjects' pronunciation time—or, more precisely, reading time—of the to-be-remembered items was measured, as a means of measuring pronunciation time differences.

Method

Subjects. Twenty-four unpaid volunteers (14 women and 10 men) from a subject pool created through local advertising participated in the experiment (mean age = 28). French was the first language of all the subjects.

Materials. The stimuli used here were developed by Richard (1993); 120 words and 120 nonwords were used to construct 24-word and 24-nonword lists. Each list comprised five items. Nonwords were consonant–vowel–consonant (CVC) trigrams, and each list of nonwords was assembled according to the construction criteria proposed by Hilgard (1951): (1) No letter appeared more than once within a list; (2) the last consonant of a nonword never immediately preceded in the alphabet the first consonant of the next nonword; and (3) nonwords were never real words in French or well-known abbreviations. Care was also taken in list construction to avoid phonological similarity, by never including rhyming items. On the other hand, words contained an average of three syllables each and they were of about equal frequency according to the Vikis–Freibergs (1974) norms. Richard (1993) reported that the mean pronunciation time for these words (545 msec) was not significantly different from that for the nonwords (548 msec).

A 2 × 2 within-subjects factorial design was used, with viewing condition (quiet vs. suppression) and lexicality (word vs. nonword) as factors. The 24-nonword lists were randomly divided into two blocks of 12 lists. The first 2 lists served as practice trials and were

not used for the analyses. Across subjects, each set was used equally often under quiet and suppression conditions. The same process was applied to the 24-word lists. The order of the four conditions was counterbalanced across subjects with a Latin square, but list order within a block was fixed.

Procedure. The subjects were tested individually within one session lasting approximately 1 h 15 min. For the immediate serial recall task, items were presented with a standard Sanyo tape recorder. Each trial began with the presentation of a 450-Hz tone lasting 500 msec. Three seconds later, items were presented at a rate of one item every 1,500 msec. The end of the lists was marked by a tone (450 Hz, 500 msec), and the experimenter stopped the tape.

Strict serial recall instructions were used. The subjects were told to recall the items in their exact order of presentation, beginning with the first one. They wrote their responses on answer sheets holding five horizontal lines numbered one to five, from left to right. They were instructed to leave a blank line, if they could not recall an item at a given serial position. They were also warned not to backtrack to change a response or fill a blank. There was no time limit for recall. The experimenter was present throughout, to ensure compliance with these instructions.

In the articulatory suppression condition, the subjects continuously repeated aloud the word *mathématiques*, at a rhythm of approximately three utterances every 2 sec. The subjects began suppressing as soon as they heard the warning tone indicating that list presentation was to begin 3 sec later, and they continued until recall was completed.

Following the immediate serial recall task, the subjects performed a reading task. They had to read aloud once all 12 lists of a block, as quickly as possible without making errors. The 12 lists of each block used in the memory task were presented in accented uppercase letters on a standard sheet of paper, with 1 list per row. Reading order of the four blocks (word and nonwords sets used under quiet and suppression conditions) was counterbalanced across subjects with a Latin square. The time needed to read each block was measured with an electronic stopwatch. Before the reading task proper, the subjects were allowed two practice trials, one with a five
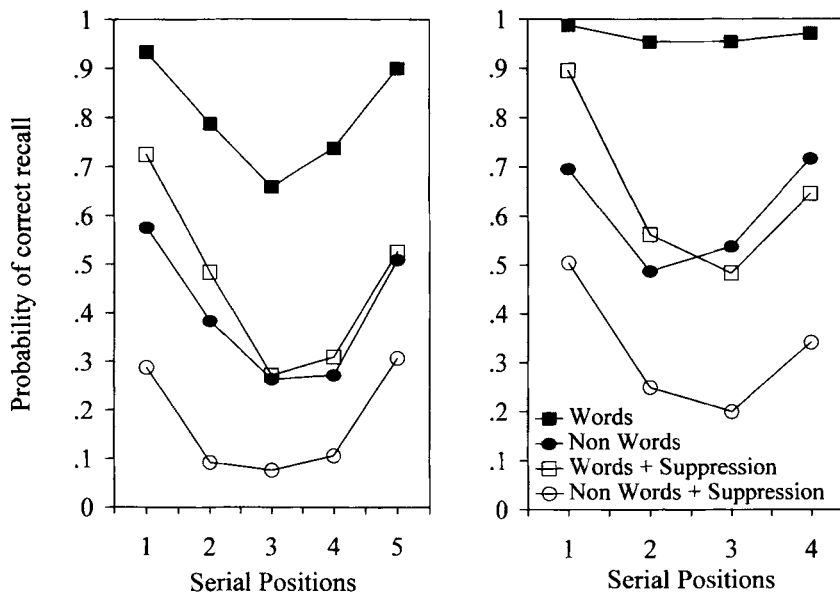

Figure 1. Mean probability of correct recall with a strict serial recall criterion as a function of lexicality, articulatory suppression, and serial position in Experiment 1 (left panel) and Experiment 2 (right panel).

nonword list and one with a list of words. The materials used for the practice trials were not the same as those presented for the memory and the reading tasks.

Results

Responses were first scored according to a strict serial recall criterion, to produce serial position curves and to facilitate comparison with previous studies. Item and order error scoring and analyses are described following this first series of results. For all analyses, unless otherwise specified, the .05 level of significance was adopted.

Strict-scoring results. The probability of correct recall as a function of lexicality and articulatory suppression is shown in the left panel of Figure 1. This figure reveals the classic serial position curve for this type of task and auditory presentation. Each serial position showed a sizable advantage of words over nonwords and a suppression impairment. Moreover, under suppression, performance in the nonword condition is extremely low: On 96.7% of the trials, two or fewer items were recalled, and on 37.2% of the trials, no response was given. By contrast, for the word suppression condition, the subjects recalled more than two items on 41.4% of the trials, and there were only 5.8% of trials in which no response was given.

A $2 \times 2 \times 5$ repeated measures analysis of variance (ANOVA) confirmed these trends, with a main effect of lexicality [$F(1,23) = 392.94, MS_e = 0.0367$], articulatory suppression [$F(1,23) = 160.95, MS_e = 0.0594$], and serial position [$F(4,92) = 38.77, MS_e = 0.0442$]. There were also three significant interactions. The higher order interaction between lexicality, suppression, and serial position was significant [$F(4,92) = 5.62, MS_e = 0.0213$]. This interaction, as well as the remaining two, is mainly attributable to a floor effect: Recall performance of nonwords under suppression was near zero for the three middle serial positions. In the same vein, the interaction between lexicality and serial position was significant [$F(4,92) = 3.34, MS_e = 0.0171$]. Simple main effects,

with an alpha level set at .01, revealed that the effect of lexicality was significant at all serial positions. Finally, there was a significant interaction between lexicality and suppression [$F(1,23) = 8.39, MS_e = 0.0464$], but simple main effects showed a significant effect of lexicality under both quiet and suppression conditions, as well as a significant effect of suppression for words and nonwords.

Error analyses. Each error was scored as either an item or an order error. An item error was defined as a missing (a blank) or a wrong item (a recalled item that had not been presented). Because presentation was auditory and because, by definition, there is no generally accepted way of writing a nonword, a lenient scoring criterion was applied. This means that all the possible spellings of a nonword were considered correct (in French, writing *rek* for *rec* or *gir* for *jir*). Similarly, all misspellings of words were accepted. In addition, misspellings of nonwords that were close to the actual sound of the presented nonword and could be considered as misheard items were counted as correct (in French, writing *gob* for *kob*, *tab* for *tad*, or *lacs* for *lax*). It should be noted that the same pattern of results was obtained when they were considered as item errors.

An order error was counted for every presented item that was recalled at the wrong serial position. However, this is not the most appropriate measure of order retention. Empirically, item and order errors are not independent: If more items are recalled, the probability of an order error is increased. For example, if no item is recalled, order errors are not a possibility, whereas if two items are recalled, the number of possible order errors is less than it would be if four items were recalled. Given that words and nonwords differ in item recall level, it is necessary to partial out the influence of item recall in the measure of order retention. To this end, proportions of order errors per item recalled were computed by dividing the total number of order errors by the total number of items recalled, regardless of the order (Murdock, 1976; Poirier

Table 1
Mean Number of Item and Mean Proportion of Order Errors per List as a Function of Lexicality, Articulatory Suppression, and Experiment

Condition	Error Type			
	Item Errors		Proportion of Order Errors	
	M	SD	M	SD
Experiment 1				
Words				
Quiet	0.66	0.57	0.08	0.06
Articulatory suppression	2.25	0.60	0.16	0.14
Nonwords				
Quiet	2.94	0.71	0.02	0.04
Articulatory suppression	4.07	0.37	0.05	0.08
Experiment 2				
Words				
Quiet	0.10	0.16	0.01	0.02
Articulatory suppression	1.23	0.66	0.07	0.06
Nonwords				
Quiet	1.55	0.64	0.01	0.03
Articulatory suppression	2.67	0.49	0.03	0.04

& Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999b). This measure of order retention ensures that, for example, with random responding the proportion of order errors would remain constant for various levels of item recall. In addition, Saint-Aubin and Poirier (1999b) showed that, contrary to the absolute number of order errors, the proportions produced the same pattern of results as that obtained with tasks that are nominally pure measures of order retention, such as the order reconstruction task and immediate serial recall with a limited pool (see Saint-Aubin & Poirier, 1999a, for an extensive discussion).

As is shown in the top section of Table 1, there are more item errors under suppression than under quiet conditions, and there are more item errors for nonwords than for words. The effect of lexicality seems stronger under quiet conditions, perhaps because of the floor effect for nonwords under suppression. A 2 × 2 repeated measures ANOVA revealed a significant effect of lexicality [$F(1,23)$ = 720.07, MS_e = 0.1391] and of articulatory suppression [$F(1,23)$ = 275.76, MS_e = 0.1609] and an interaction [$F(1,23)$ = 5.18, MS_e = 0.2384]. Simple main effects, with an alpha level set at .01, revealed that the effect of lexicality was significant under both quiet and suppression conditions and that the effect of suppression was significant for both words and nonwords. The ω^2 (for fixed-effects model) statistic indicated that lexicality accounted for 56.6% of the variance, suppression for 25.0%, and the interaction for 0.6%.

With proportions of order errors per item recalled, presented in the top section of Table 1, there are fewer errors for nonwords, and articulatory suppression is associated with a greater proportion of order errors. The repeated measures ANOVA showed an effect of lexicality [$F(1,23)$ = 20.15, MS_e = 0.0079] and articulatory suppression [$F(1,23)$ = 7.28, MS_e = 0.0093], but the interaction failed to reach significance (F = 1.50). However, despite the nonsignificant interaction, the effect of suppression on nonwords might be more apparent than real, because there were practically no order errors for nonwords: On a total of 1 200 presented nonwords per condition (24 subjects × 10 lists × 5 item per list) there were 16 nonwords recalled at the wrong serial position under suppression, and 13 in the quiet condition. The ω^2 showed that lexicality accounted for 15.3% of the variance, suppression for 5.9%, and the nonsignificant interaction for 0.4%.

A final analysis was carried out to compare the mean reading times for words and nonwords. The repeated measures ANOVA revealed that the mean reading time for a single recitation of a five-word list (2.32 sec) was faster than for a list of nonwords [2.69 sec; $F(1,23)$ = 24.60, MS_e = 0.0698]. This effect accounted for 19.5% of the variance (ω^2).

Discussion

As in previous studies, with a strict serial recall criterion, words are better recalled than nonwords (Besner & Davelaar, 1982; Hulme et al., 1991; Hulme et al., 1995; Richard, 1993). This effect is attributable to the better item recall of words, an effect that clearly overshadows the order recall advantage found for nonwords. Articulatory suppression hindered item and order information recall for words and nonwords, but there are uncertainties concerning its effect on order recall for nonwords, given the very small number of order errors on which it is based.

Results also showed that words are pronounced faster than nonwords. This further illustrates the difficulty of equating words and nonwords on pronunciation duration, because with the exact same stimuli and procedure, Richard (1993) found no significant difference. Although speculative, one possibility could be that the overall faster pronunciation of the subjects in the present study—501 msec per item versus 547 msec for Richard—affects more nonwords by provoking, for example, more hesitations. However, because articulatory suppression is known to abolish the effects of pronunciation rate—under suppression, recall is similar for items of various pronunciation duration (Baddeley et al., 1984)—it seems safe to conclude that lexicality has an effect over and above any confound with pronunciation rate.

EXPERIMENT 2

In Experiment 1, recall performance for nonwords under suppression was very low. This was predicted by both the retrieval-based hypothesis and general accounts of residual performance as a contribution of long-term knowledge (see, e.g., Hulme et al., 1991; Zhang & Simon, 1985). However, although expected, the very low recall performance might raise problems for interpreting the marginal impact of articulatory suppression on order information recall for nonwords. It can be argued that this situation arises because of a floor effect. In addition, as will be seen in the General Discussion section, Burgess and Hitch's (1992, 1996) connectionist model of short-term memory might account for the opposite effects of lexicality on item and order information recall under quiet but not under suppression conditions. Consequently, it is particularly important to document the effects of lexicality on order information under suppression when a floor effect is unlikely. Accordingly, Experiment 2 replicated Experiment 1 with shorter lists holding four items instead of five. It should be noted that by reducing list length, recall performance of words under quiet conditions is likely to be at ceiling. However, this would not raise problems, because the impact of lexicality under quiet conditions has already been assessed in Experiment 1, where a ceiling effect was unlikely, and because the main aim of Experiment 2 was to evaluate the impact of articulatory suppression on order information recall for nonwords and the impact of lexicality under articulatory suppression.

In sum, both Experiments 1 and 2 were needed to fully examine the impact of lexicality on item and order information, under quiet and suppression conditions. This situation arose because both lexicality and suppression produced large effects. Consequently, if under the quiet condition, word data were off the ceiling—as in Experi-

ment 1—nonword data tended to be on the floor in the suppression condition; and if nonword data were off the floor, word data tended to be at ceiling in the quiet condition. Together, Experiments 1 and 2 would provide answers that neither one could provide in isolation.

Method

Subjects. Twenty-four unpaid volunteers (17 women and 7 men) from a subject pool created through local advertising participated in the experiment (mean age = 22). French was the first language of all the subjects, and none of them had participated in the previous experiment.

Materials and Procedure. The materials for this experiment were constructed by removing one item from each list of words and nonwords used in Experiment 1. Care was taken to ensure that the new version of the lists of nonwords respected the construction criteria proposed by Hilgard (1951). In addition, one nonword was replaced because it was the name of a new company in the area. However, by inadvertence, the replacing nonword rhymed with another nonword within the list. In sum, two sets with 12 four-nonword lists and two sets with 12 four-word lists were used—with the first 2 lists of each trial serving as practice trials. The procedure for the memory task was exactly the same as that in Experiment 1.

The reading task was slightly modified. Instead of reading all 10 lists of a block once within the same trial, a trial now consisted of a single list read five times. This form of reading task is more common in the field and uses list lengths that are more typical of immediate serial recall (see, e.g., Hulme et al., 1991). The subjects read either 2 lists of words and 3 lists of nonwords or 3 lists of words and 2 lists of nonwords. The trials were blocked by lexical status, and the reading order of the two blocks was completely counterbalanced across subjects. Lists were assigned to subjects on a random basis, with the constraint that over the 24 subjects each list has been presented to 3 of them.

Results

As in the previous experiment, performance was scored according to a strict serial recall criterion, and item and order errors were analyzed separately. Strict scoring results are presented in the right panel of Figure 1, showing for all conditions a higher recall level than that in Experiment 1. Except for words under the quiet condition, where performance was at ceiling, the usual serial position curve was found, and at each serial position, there was a large advantage of words over nonwords and a strong suppression impairment. A $2 \times 2 \times 4$ repeated measures ANOVA confirmed these trends, with a main effect of lexicality [$F(1,23) = 208.92$, $MS_e = 0.0532$], articulatory suppression [$F(1,23) = 343.43$, $MS_e = 0.0256$], and serial position [$F(3,69) = 61.30$, $MS_e = 0.0172$]. There were also three significant interactions. The higher order interaction between lexicality, suppression, and serial position was significant [$F(3,69) = 6.58$, $MS_e = 0.0150$]. This interaction, as well as the remaining two, is mainly attributable to a ceiling effect: Recall performance is almost perfect at all serial positions for words under the quiet condition. Similarly, the interaction between lexicality and serial position [$F(3,69) = 2.90$, $MS_e = 0.0173$], as well as the interaction between suppression and serial position [$F(3,69) = 17.88$, $MS_e = 0.0173$], was significant. Simple main effects, with an alpha level set at .01,

revealed that the effects of both lexicality and suppression were significant at all serial positions.

Item error results. Examination of the bottom section of Table 1 reveals a greater number of item errors for nonwords than for words under both the quiet and the suppression conditions and increased errors for both types of item under suppression. A repeated measures ANOVA, with lexicality and suppression as factors, confirmed these trends, with a main effect of lexicality [$F(1,23) = 271.91$, $MS_e = 0.1840$] and of suppression [$F(1,23) = 279.38$, $MS_e = 0.1091$] and a nonsignificant interaction ($F < 1$). The ω^2 statistic indicated that lexicality accounted for 47.0% of the variance and suppression for 28.7%; it was impossible to compute this statistic for the interaction, given that the F ratio was smaller than 1.

Order error results. With proportion of order errors per item recalled, presented in the bottom section of Table 1, there were fewer errors for nonwords than for words under suppression but not under quiet conditions, owing to the ceiling effect. Suppression was associated with a greater proportion of order errors for words and, to a lesser extent, for nonwords. The repeated measures ANOVA confirmed these trends, with a main effect of lexicality [$F(1,23) = 7.01, MS_e = 0.0017$] and of suppression [$F(1,23) = 19.32, MS_e = 0.0020$] and an interaction [$F(1,23) = 7.61, MS_e = 0.0013$]. Simple main effects revealed that the effect of lexicality was significant only under suppression [$F(1,46) = 14.49, MS_e = 0.0015$] and that the effect of suppression was significant for words [$F(1,46) = 26.57, MS_e = 0.0016$], but not for nonwords [$F(1,46) = 2.89, p = .10$]. The ω^2 showed that lexicality accounted for 4.5% of the variance, suppression for 15.6%, and the interaction for 3.6%.

Finally, mean reading times for words and nonwords were compared. The repeated measures ANOVA revealed that the mean reading times for a single recitation of a four-word list (1.56 sec) and a four-nonword list (1.60 sec) were not significantly different ($F < 1$).

GENERAL DISCUSSION

Considered together, the results of the experiments replicated the usual better recall of words over nonwords when performance is assessed with a strict scoring criterion (Besner & Davelaar, 1982; Hulme et al., 1991; Hulme et al., 1995; Richard, 1993). More important, to the best of our knowledge, this is the first demonstration that the overall advantage of words over nonwords is attributable to their better item recall, which masks the better order recall of nonwords. This pattern of results has been observed under both quiet and suppression conditions. Suppression was found to increase item errors for both words and nonwords, but its detrimental effect on order information was, for all practical purposes, restricted to words. Finally, the effects found under suppression—a condition known to abolish the word length effect—and the reading time results of Experiment 2 indicate that the

present results are unlikely to be attributable to a confounded effect of articulation speed.

The Retrieval-Based Hypothesis

The retrieval-based hypothesis presented in the introduction can easily account for the effects of lexicality and suppression—as well as those of frequency and semantic similarity—on item and order information. According to this proposal, at recall, degraded phonological representations set up by list presentation are output in the appropriate order and subject to a reconstruction process calling on long-term knowledge of the to-be-remembered items (Brown & Hulme, 1992, 1995, 1996; Hulme et al., 1991; Hulme et al., 1995; Poirier & Saint-Aubin, 1995, 1996; Roodenrys et al., 1994; Saint-Aubin & Poirier, 1999b; Schweickert, 1993). The greater number of item errors for nonwords observed in the present study can be accounted for by calling on their inadequate long-term representations, preventing reconstruction of their degraded representations. A similar pattern of item errors was found with word frequency and semantic similarity: More item errors for less frequent and semantically dissimilar items (Crowder, 1979; Murdock, 1976; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999b). Long-term representations are thought to be easier to access for frequent words and semantically similar items. On the other hand, the greater number of item errors under suppression is attributed to the greater degradation of phonological traces (Baddeley, 1990; Nairne, 1990; Neath & Nairne, 1995), because the retrieval process, based on the remaining phonological representations, is thought to be less efficient.

As for item errors, order errors are also attributed to trace reconstruction problems during the retrieval process (for related proposals, see Lewandowsky & Murdock, 1989; Nairne, 1990; Schweickert et al., 1990). More precisely, it is assumed that, at recall, degraded phonological traces are output in the appropriate order. Order errors occur because it is not always possible to uniquely identify one of the list items on the basis of the degraded phonological traces. This happens because phonological features are likely to be embodied in a number of list items. Thus, if a phonological trace has lost most of its distinctive features, it is easy to see how the long-term representation of another list item, holding common features, can be erroneously selected as an appropriate recall candidate (Schweickert et al., 1990; Sperling & Speelman, 1970). This is reminiscent of the phonological similarity effect. Within similar lists, to-be-remembered items share more phonological features than do items in dissimilar lists. This would entail a higher probability, at recall, of a degraded phonological trace holding shared phonological features. This would translate into a higher probability of order errors. This can also account for the greater proportion of order errors for words under suppression. By increasing degradation, suppression would also increase the probability of losing distinctive phonological features. With less distinctive features, a higher probability

of order errors is anticipated for words because, for those items, the retrieval process is operating, given that words can take advantage of useful long-term representations.

The retrieval-based hypothesis also provides a straightforward account of the very low proportions of order errors found for nonwords, resulting in better recall of order information for nonwords than for words under both quiet and suppression conditions. Without adequate long-term representations, the retrieval process is not effective for nonwords. Because order errors are attributed to the retrieval process, the lack of adequate long-term representations for the unfamiliar nonwords used here should translate into very few order errors. In the same vein, articulatory suppression should not influence the proportions of order errors for nonwords. This would happen because, even if, by increasing degradation, suppression increases the probability of losing distinctive phonological features, it would not influence the proportion of order errors, because the reconstruction process is inoperative for nonwords.

It should be noted that it has been suggested that reconstruction might still be possible for nonwords on the basis of long-term phonological information (Schweickert, 1993). For example, recall performance of nonwords under suppression has been investigated in another study, where, however, performance was not assessed separately for item and order information (Besner & Davelaar, 1982). In the Besner and Davelaar study, recall of nonwords under suppression was quite good, varying, across experiments, between 38% and 45% correct. However, the nonwords used by Besner and Davelaar were wordlike (e.g., *snude, zede, smude*). Consequently, long-term phonemic information might have been used to reconstruct the degraded phonological traces. However, because the nonwords used here had very low wordlikeness, the long-term phonemic information would have allowed very limited reconstruction opportunities, explaining the very limited number of order errors.

Alternative Accounts

A number of other models can also account for the influence of lexicality on immediate serial recall performance, but only when it is assessed with an overall performance measure factoring in item and order information (see, e.g., Brown & Hulme, 1995, 1996; Burgess & Hitch, 1996; Hartley & Houghton, 1996; Schweickert, 1993). Given that our data reveal that lexicality produces opposite effects on item and order information, a more detailed account is needed. To the best of our knowledge, the aforementioned proposals have not been designed to make distinct predictions for item and order information recall. However, such predictions can be derived from the connectionist model of short-term memory proposed by Burgess and Hitch (1992, 1996).

In Burgess and Hitch's (1992, 1996) model, the presentation of an item activates the representation of the item itself and, to a lesser extent, the representation of all items sharing its phonological features. The order of a

sequence is maintained both by associating a different temporal context to each item (position–item association) and by associating the phonemes of an item to the phonemes of the next item (chaining). According to this view, there are fewer item errors for words than for nonwords because, at recall or during rehearsal, words activate their constituent phonemes more strongly, owing to long-term connections (Burgess & Hitch, 1996). The smaller proportion of order errors for nonwords can be accounted for by the fact that the words used here were longer than the nonwords—in order to obtain similar pronunciation durations. Being longer, words within a list are more likely to share phonological features than are nonwords, even if care was taken to avoid rhyming items. Common phonological features hurt order recall through the phonemic activity. Articulatory suppression is implemented by preventing the use of chaining weights, resulting in virtual inactivity at the phoneme level; at the behavioral level, the phonological similarity effect disappears (Burgess & Hitch, 1992). Consequently, order recall should be similar for words and nonwords under suppression, because serial order is maintained solely by the context–item associations that are not affected by lexicality.

Conclusion

In sum, Burgess and Hitch's (1992, 1996) model can make distinct predictions concerning the influence of lexicality on item and order recall. It can successfully account for the effect of lexicality on item recall and for the better order recall of nonwords under the quiet condition, but not under suppression. This is problematic, because our results show that, under suppression, there is still a clear disadvantage for words with respect to order information recall. By contrast, the retrieval-based hypothesis predicted the effects of lexicality and articulatory suppression on both item and order information. This proposal can also account for the influence of the two other long-term memory factors with known effects on item and order information in immediate serial recall—namely, word frequency and semantic similarity. Finally, the results of Walker and Hulme (1999) also support the predictions of the retrieval-based hypothesis for imagery: an effect on item recall and no effect on order recall. This implies that this framework can provide a general account of the detailed effects on item and order information of all the long-term memory factors that have been investigated to date.

REFERENCES

BADDELEY, A. D. (1990). *Human memory: Theory and practice.* Needham Heights, MA: Allyn & Bacon.

BADDELEY, A. D., LEWIS, V. J., & VALLAR, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36A, 233-252.

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

BESNER, D., & DAVELAAR, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology*, 36, 701-711.

BROWN, G. D. A., & HULME, C. (1992). Cognitive psychology and second language processing: The role of short-term memory. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 105-122). Amsterdam: North-Holland.

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

BROWN, G. D. A., & HULME, C. (1996). Nonword repetition, STM, and word age-of-acquisition: A computional model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 129-148). Hove, U.K.: Psychology Press.

BURGESS, N., & HITCH, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory & Language*, 31, 429-460.

BURGESS, N., & HITCH, G. J. (1996). A connectionist model of STM for serial order. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 51-72). Hove, U.K.: Psychology Press.

CROWDER, R. G. (1979). Similarity and order in memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 13, pp. 319-353). New York: Academic Press.

HARTLEY, T., & HOUGHTON, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory & Language*, 35, 1-31.

HILGARD, E. R. (1951). Methods and procedure in study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 517-567). New York: Wiley.

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 & Language*, 30, 685-701.

HULME, C., ROODENRYS, S., BROWN, G., & MERCER, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, 86, 527-536.

HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D. A., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 1217-1232.

LEWANDOWSKY, S., & MURDOCK, B. B. (1989). Memory for serial order. *Psychological Review*, 96, 25-57.

MURDOCK, B. B., JR. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, 105, 191-216.

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

NEATH, I., & NAIRNE, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, 2, 429-441.

POIRIER, M., & SAINT-AUBIN, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology*, 48A, 384-404.

POIRIER, M., & SAINT-AUBIN, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, 50, 408-412.

RICHARD, P. A. (1993). *Contribution de la mémoire à long terme lors du rappel sériel immédiat dans la perspective du modèle de Baddeley* [Long-term memory contribution to immediate serial recall performance within Baddeley's model]. Unpublished master's thesis, Université Laval.

ROODENRYS, S., HULME, C., ALBAN, J., ELLIS, A. W., & BROWN, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition*, 22, 695-701.

SAINT-AUBIN, J., & POIRIER, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology*, 34, 347-352.

SAINT-AUBIN, J., & POIRIER, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology*, 52A, 367-394.

SCHWEICKERT, R. (1993). A multinominal processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, 21, 168-175.

SCHWEICKERT, R., & BORUFF, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 12, 491-425.

SCHWEICKERT, R., GUENTERT, L., & HERSBERGER, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, 1, 74-77.

SPERLING, G., & SPEELMAN, R. G. (1970). Acoustic similarity and auditory short-term memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 151-202). New York: Academic Press.

VIKIS-FREIBERGS, V. (1974). *Fréquence d'usage des mots au Québec* [Frequency of usage of French words in Quebec]. Montréal: Les presses de l'Université de Montréal.

WALKER, I., & HULME, C. (1999). Concrete words are easier to recall than abstract: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1256-1271.

ZHANG, G., & SIMON, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, 13, 193-201.