Language Learning 53:1, March 2003, pp. 67–121

Phonological Memory and Rule Learning

John N. Williams and Peter Lovatt
University of Cambridge

Two experiments examined the relationship between individual differences in phonological memory (PM) and the ability to learn determiner-noun agreement rules in semiartificial microlanguages. Participants were tested on their ability to induce the grammatical gender of nouns from the distribution of the determiners that accompanied them. Three measures of PM were found to be related to rule learning as assessed by a generalization test: phonological short-term memory, vocabulary learning, and memory for determiner-noun combinations early in the experiment. There were also statistically independent effects of knowledge of other gender languages, suggesting that both memory and nonmemory factors were related to learning outcomes.

Recent work within a cognitive approach to first language (L1) and second language (L2) acquisition has stressed the role of phonological memory (PM). For example, Baddeley, Gathercole, and Papagno (1998) argue that phonological short-term memory (PSTM, or more specifically the "phonological loop") is a

---

John N. Williams and Peter Lovatt, Research Centre for English and Applied Linguistics.

Many thanks to the Language Learning reviewers for their valuable comments.

Correspondence concerning this article should be addressed to John N. Williams, Research Centre for English and Applied Linguistics, University of Cambridge, Keynes House, Trumpington Street, Cambridge, CB2 1QA, UK. Internet: jnw12@cam.ac.uk

"language learning device" and that it is involved in learning vocabulary and grammar in the L1 and L2. The goal of the present research was to evaluate this claim experimentally by examining the relationship between measures of PM and the ability to learn grammatical rules.

PSTM is held to underlie people's ability to immediately repeat novel phonological sequences (for example, lists of random digits, random words, nonsense words) or simply to repeat single nonwords. Performance on these tasks depends upon the ability to maintain representations in PM for just long enough to support recall before they are lost, through either a process of decay or interference from other phonological material (Baddeley, 1990). Although the time over which material in PSTM is retained may be enhanced through rehearsal strategies, the amount of material that can be retained is assumed to reflect the "capacity" of PSTM itself.[1]

PSTM is just one component of what Baddeley (1990) refers to as the "working memory" system. The term "working memory" is also widely used to refer to the total resources that are available to the individual for simultaneous processing and storage (Just & Carpenter, 1992). According to this view the individual possesses finite resources that are consumed by both the processing and storage of information. This means that the processing and storage demands of a task can be traded off against each other. For example, in an easy or well-practiced task, processing demands will be low and so storage capacity will be relatively high. In this view it makes little sense to measure the "storage capacity" of the individual without reference to a particular processing task, and tests of "working memory" should involve simultaneous storage and processing (Daneman & Carpenter, 1980; Engle, Kane, & Tuholski, 1999). However, there is currently a good deal of debate over whether it is correct to think of processing and storage as being related by a central resource pool. Engle, Tuholski, Laughlin, and Conway (1999) provide evidence that individual differences on a variety of verbal and nonverbal tasks are best explained by a model comprising domain-specific storage capacities plus a domain-general attentional resource that is related to general fluid intelligence. This attentional component is similar to Baddeley's (1990) notion of the "central executive": an attentional system that is involved in regulating information processing and that is distinct from the specific storage systems used to maintain information during task performance. These more complex notions of working memory are beginning to be considered in relation to language learning aptitude (Sawyer & Ranta, 2001), and some empirical evidence relating them to language learning success is beginning to emerge (Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Robinson, 2002). However, here we chose to focus on the issue of storage, since all models of working memory assign some degree of importance to the efficiency of domain-specific storage systems, whether or not this impinges on the resources available for other cognitive operations.

It is now well documented that individual differences on tasks that tap PSTM ability are related to vocabulary learning in both normal and language-impaired adults and children in L1 and L2 acquisition. One often-cited example is an Italian patient PV (Baddeley, Papagno, & Vallar, 1988), whose ability to repeat lists of random digits/words was minimal and who could not learn pairs of L2 (Russian) words and L1 (Italian) translations. In contrast, she could learn novel pairs of words in her native Italian, and her comprehension of Italian appeared to be unimpaired. Although this pattern casts doubt on the role of PSTM in processing and learning material in a known language, it does suggest that PSTM plays an important role in the formation of phonological representations of novel forms. In normal child L1 acquisition, Gathercole, Willis, Emslie, and Baddeley (1992) found a relationship between PSTM at age 4 (as measured by nonword repetition) and vocabulary size a year later. In adult L2 acquisition, Service and Craik (1993) demonstrated a relationship between PSTM and vocabulary learning efficiency.

It may seem counterintuitive that there should be a relationship between the ability to immediately repeat novel phonological forms and vocabulary learning. One is a measure of PSTM, whereas the other is presumably a reflection of phonological long-term memory (PLTM). Yet, as Gathercole and Martin (1996) note, both short-term and long-term retention of phonological forms are affected by the same factors, such as word length, phonological similarity, and articulatory suppression. This implies that the two storage systems are closely related. Gathercole and Martin propose that temporary representations in PSTM are used to build PLTM structures and that these two aspects of PM are in constant interaction.

Given the close relationship between PSTM and PLTM, we shall refer to a more general PM ability to refer to individual differences in the ability to store phonological forms in the short, medium, or long term. Whereas the link between PM ability and vocabulary learning has been the focus of earlier work, here we focus on its relationship to grammar learning.

"Emergentist" approaches to language learning (Ellis, 1998) posit a close relationship between grammar and the lexicon and therefore would seem to predict that any factor that influences vocabulary learning should also influence grammar learning. For example, in construction grammar (Goldberg, 1998), words, bound morphemes, and syntactic structures are all represented as "items" in the lexicon over which abstraction processes operate. The usage-based account of L1 acquisition (Tomasello, 2000) stresses simple imitative learning of adult utterances. Abstract categories and schemas emerge from representations of these utterances in memory. Ellis (1996) argues that the storage of morphemes and morpheme sequences conforms to the same principles as memory for sequences of phonemes within words; the relevant chunks are just larger. Grammatical rules emerge through a chunking mechanism that operates on long-term memory representations of morpheme sequences in the same way that lexical representations emerge from long-term memory representations of phoneme sequences. Whether the learning process is implicit or explicit, "one way or another, a knowledge of grammar comes from analysis of a large stock of learned exemplar sequences" (Ellis, 1996, p. 115). Some connectionist models of language learning have stressed learning sequence information as the basis not only for learning lexical items (Saffran, Newport, & Aslin, 1996), but also for inducing noun classes (Elman, 1990; Redington & Chater, 1998). All of the above can be regarded as "memory-driven" accounts of learning because they stress the role of long-term memories for morpheme combinations as the basis for grammatical development. Grammatical rules emerge as generalizations over representations of morpheme sequences, and these sequences are stored in the same memory system as the morphemes themselves. Because PM relates to vocabulary development, it should also be related to grammatical development.

Even though the above line of argument is intuitively plausible, it does not necessarily follow that, just because PM is related to the storage of vocabulary, it is also related to the storage of sequences of morphemes. The relationship between performance on PSTM tasks and vocabulary learning is assumed to result from the function of PSTM in maintaining representations of novel phonological forms (Baddeley et al., 1998). But it is not unreasonable to suppose that many grammatical rules are learned after the relevant morphemes themselves have been segmented out of the input and become familiar as discrete lexical units. For example, articles and nouns may be familiar as segmented form-level units, but the learner has to remember combinations of articles and nouns and the contexts in which they occurred to discover the rules governing article usage. In such cases it is the storage of sequences of familiar morphemes that is critical to grammar learning, rather than the storage of unfamiliar phonological forms. Although in L1 acquisition it may be difficult to separate the emergence of morphological units and the analysis of their distributional properties in all cases, in the case of adult L2 acquisition it will tend to be the case that the individual morphemes are familiar because they have been learned as individual items, and it is their sequencing that needs to be analyzed in the course of grammar learning. It needs to be established that the ability to remember sequences of *familiar* morphemes over intervals sufficient to support grammar learning is related to PM ability.

Memory research is traditionally concerned with the short- and long-term retention of material composed of familiar units, such as random-word or -digit lists, sentences, and texts. It is interesting that this research would lead one to be skeptical that the medium for storing sequences of familiar units would be phonological. Early work on sentence memory stressed the transience of memory for exact surface form. Information about exact wording and word order is assumed to be lost as soon as the underlying meaning of an utterance has been computed (Gernsbacher, 1990; Jarvella, 1971). In this view, word order information is retained in PSTM for only 2 s or so. Longer term memory would be dependent upon another form of representation, which in the case of sentence comprehension is assumed to be either in a propositional (Kintsch, 1988) or situation/mental model (Johnson-Laird, 1983) format. Longer term retention is therefore thought to be subject to meaning-level factors such as propositional cohesion (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975) and semantic elaboration (Craik & Lockhart, 1972), rather than PM. In fact, PV (Baddeley et al., 1988) provides a striking illustration of this distinction. As already mentioned, her inability to associate novel words with L1 words (as an exercise in foreign vocabulary learning) is attributed to her poor PM ability. Yet when she was required to learn arbitrary associations between pairs of L1 words (paired-associate learning), her performance was unimpaired. Clearly, learning associations between familiar morphemes is not necessarily dependent on PM.

Theoretical models of PSTM also seem to imply that longer term retention of sequences of familiar morphemes should be unrelated to PM ability. There are a number of models of PSTM that seek to explain how the ordering of familiar items is retained long enough to support immediate recall (Page & Henson, 2001). None of these models explain how order information might be retained in long-term memory. Yet it is known that even the order of random-word lists can be retained at well above chance levels at recall intervals that are beyond the duration of PSTM, even when rehearsal is prevented (see the studies described in Page & Henson, 2001). Page and Henson (2001) suggest that this ability may depend upon a nonphonological form of representation that they dub the "positional-episodic record" (p. 184). This is similar to Baddeley's (2000) notion of the "episodic buffer": an intermediate-term storage system that integrates information from different modalities (e.g., phonology and semantics) and stores it using a "common multi-dimensional code" (p. 421) before it is transferred to long-term memory.[2]

Given the above evidence, one might argue that if grammatical rules are to emerge from representations of input in long-term memory, then those representations are unlikely to be of a phonological form. Yet there are cases of linguistic rules that do depend upon phonological distinctions. For example, in English only forms consisting of a phonological foot can in principle take both prepositional and double-object datives, although semantic properties of the verb are also a factor (Pinker, 1989). Or in the case of languages involving grammatical gender, the learner must come to realize that although certain forms, such as determiners, might mean the same thing and have similar grammatical functions, their precise form depends upon the gender of the accompanying noun. For example, in Italian, *il* and *la* are both singular definite articles. Yet the learner must remember which of these occurs with which nouns if the underlying noun class distinction is to be learned. This would appear to require retention of the phonological form of the determiner, even though its meaning and syntactic properties are all that is required for successful comprehension. In general, although rapid loss of phonological information might apply to comprehension of a known language, in the case of a developing language system, it is not so clear that learners can afford to be quite so disinterested in phonological form, lest they "throw out the baby with the bathwater."

What is the empirical evidence for the relationship between PM and grammar learning? In the case of L1 acquisition, Bates and Goodman (1997) review an impressive range of studies that show that there is indeed a consistent relationship between vocabulary size and syntactic development in both normal and abnormal populations. One way of interpreting this would be that, since vocabulary development is an indicator of PM ability, PM must also be related to grammar learning. However, alternative explanations are possible. Bates and Goodman themselves suggest that the vocabulary-grammar relationship could exist because vocabulary knowledge influences the segmentability and salience of grammatical morphemes ("perceptual bootstrapping"). But this means that grammar learning could be a result of a non-memory-based process (such as parameter setting in a principles and parameters account), a process that is merely dependent upon vocabulary learning to deliver discrete morphemic units.

Adams and Willis (2001) argue that PSTM is not related to L1 grammar learning. Although measures of PSTM do relate to the complexity of children's speech (Adams & Gathercole, 1996), Adams and Willis argue that this is because of the role of PSTM in the speech production process. Individuals with greater phonological working memory will be able to produce sentences with more complex syntax, but this does not imply any relationship between PM and the acquisition of grammatical knowledge. Indeed, there is very little evidence for a relationship between PSTM and comprehension ability. Speidel (1993) studied a child who had poor PSTM and grammatically impoverished spoken production but who also showed no signs of a deficit in comprehension. Adams, Bourke, and Willis (1999) report weak correlations between PSTM and comprehension, but these are accounted for by an intercorrelation with verbal fluency, which has been associated with the efficiency of the attentional mechanism.

In L2 acquisition Skehan (1986, 1998) provides evidence for a relationship between a memory component of language aptitude and L2 learning success. However, he suggests that PM ability is involved in the storage of phrases that are then employed in production as unanalyzed units. No specific claims are made concerning the relationship between PM ability and grammar learning.

More positive evidence for a relationship between PM and grammar learning has come from experimental studies. Ellis and Sinclair (1996) compared learning in two groups of participants who were exposed to single words and phrases in Welsh. One group repeated each phrase as it was presented, whereas the other group was prevented from repeating by articulatory suppression (continuously counting from one to five). On nearly all measures of vocabulary and grammar learning the repetition group was superior. This demonstrates that learning is dependent on the formation of phonological representations of input sequences, a process that is disrupted by articulatory suppression. Ellis, Lee, and Reber (1999) provide evidence that learning a local-agreement rule is related to PM ability. During the learning phase participants were simply required to recall strings of nonsense words immediately after they were presented (e.g., *miu-ra ko-gi pye-ri lon-da*, in which the middle two words exhibit local agreement). It was found that the accuracy of recall for the first few input strings correlated with performance on tests of rule knowledge at the end of the learning phase. Ellis et al. assumed that immediate recall of input at the beginning of the experiment is equivalent to a PSTM test for novel phonological forms. They therefore concluded that PSTM ability is related to rule learning. However, in both of these studies the morphemes were all unfamiliar at the start of training. PSTM would be expected to be related to the rate at which representations of the individual morphemes are established in PLTM. But as noted above, it is possible that, once the morphemes are familiar, retention of sequencing information is not influenced by PM ability. Alternatively, the more familiar the morphemes become, the more salient they are to other, non-memory-based learning processes that extract the sequencing regularities. It is posible, therefore, that in these experiments PM affected grammar learning only via its effect on vocabulary learning.

Williams (1999) examined incidental learning of some aspects of Italian grammar. As in Ellis et al. (1999), some of the target rules concerned agreement (e.g., between nouns and determiners and adjectives). Participants performed a memory task on input sentences, although recall was after one intervening item, unlike in Ellis et al., where immediate recall was required. Here too memory performance on the first few input sentences correlated with rule learning. There were also correlations between rule learning and vocabulary learning efficiency. Assuming that delayed recall of input and vocabulary learning reflect PM ability, then it can be inferred that PM ability affects rule learning. This result is less likely to be due to the effect of PM on vocabulary learning than in Ellis and Sinclair (1996) and Ellis et al. (1999), because the participants were pretrained on the content words used in the experiment. However, not all of the grammatical morphemes were pretrained, and so it remains possible that it was the rate at which these were learned that was critical.

A study by Robinson (1997) provides some evidence that PM might be involved in grammar learning even when the morphemes are familiar. Robinson investigated the role of various components of language learning aptitude in the learning of syntactic rules of English by intermediate-level learners of English. Learners were exposed to sentences containing target rules under four conditions: implicit (a memory task), incidental (a comprehension task), rule discovery, and instructed. Performance on a subsequent grammaticality judgment task was evaluated in relation to two components of language aptitude: memory and grammatical sensitivity. It was found that performance in the rule search and instructed conditions correlated with both grammatical sensitivity and memory, whereas in the implicit condition, learning correlated only with grammatical sensitivity (there were no correlations in the incidental condition). These results indicate a role for both memory and nonmemory factors in explicit learning. However, there was no test of participants' medium-/long-term memory for input. It needs to be established that the link between PM and rule learning is mediated by input memory itself, since it is this that is assumed to be directly causal in rule learning according to memory-driven views of learning.

The aim of the present experiments was to further explore the relationship between PM and grammar learning in adults. This issue was investigated in the context of a noun class induction problem. Different target languages were used in Experiments 1 and 2, but both languages contained noun class systems in which a noun's grammatical gender conditioned which determiners it could occur with. There were no semantic cues to gender, and so in order to discover why sometimes one determiner was used and sometimes another, it was necessary to remember the forms of the determiners that were used with specific nouns. This kind of memory-intensive learning problem is therefore an ideal starting point for examining the relationship between PM and grammar learning. Such a problem also permits relatively clear tests of generalization, that is, the ability to produce novel determiner-noun combinations on the basis of a noun's word class.

Both experiments had the following general structure: After performing an initial short-term memory task, the participants learned all of the nouns and determiners that would be used in the experiment as isolated vocabulary items. In the subsequent training task they performed what, on the surface, appeared to be a simple rote memory task for determiner-noun combinations (noun phrases). Noun phrases were presented in sets of three (Experiment 1) or four (Experiment 2), and after each set the participants had to attempt to recall each phrase. After performing this task the participants completed a generalization test in which they had to produce determiner-noun combinations that had never been encountered during training. This provided a measure of rule learning.

The following variables were of interest:

PSTM. The initial PSTM test measured the participants' short-term memory for novel phonological forms (these were actually the nouns from the target language).

Vocabulary. The vocabulary learning phase measured the efficiency with which the participants could establish representations of novel phonological forms in PLTM.

Input Memory. Performance on the training task could be used as a measure of medium-term memory for sequences of familiar morphemes, that is, as a test of the ability to establish representations of morpheme sequences in PLTM. We refer to the recall task as " medium term " because recall was required after presentation/recall of as many as four other items and could not therefore be supported by representations in PSTM (which are assumed to have a duration of only about 2 s; Baddeley, 1990). We therefore assume that recall was supported by representations in long-term memory, even though these representations would probably not have been strong enough to support recall at very long intervals. Because memory for input can obviously be influenced by rule learning, this measure has to be taken early in the training phase.

Generalization. Since the generalization test involved items that had not been presented during training, performance reflected the participants' level of learning the rules underlying the target system.

Language Background. Given the nature of the target rules, it seemed likely that knowledge of other gender languages would be an additional determinant of learning. A questionnaire was used to obtain this information.

The first set of research questions concerns the interrelationships between the memory measures PSTM, Vocabulary, and Input Memory. On the basis of previous research one would expect a relationship between PSTM and Vocabulary, since, as argued above, these two measures can essentially be regarded as a reflection of PM ability. But will PM, as operationally defined by these two measures, be related to memory for sequences of familiar morphemes as measured by Input Memory? If so it could be argued that the memory system that supports vocabulary learning is similar to that which supports memory for input sequences even when the individual morphemes are familiar and that both memory systems employ phonological representations.

The second set of research questions concerns the relationships between the memory measures and generalization test performance. First, will Input Memory be related to Generalization ? If it turns out to be, this would suggest that rule knowledge emerges from representations of the input. But this leaves open the issue of the nature of that representation. If PM (as measured by PSTM and Vocabulary) is related to Input Memory, then, as stated above, it could be concluded that the relevant representations are phonological. If Input Memory is related to Generalization, then one might also expect to find direct relationships between measures of PM (PSTM and Vocabulary) and Generalization. Note, however, that this third relationship does not necessarily follow from the existence of the other two. It is possible that the relationship between PM and Input Memory is due to one factor (phonological storage), whereas that between Input Memory and Generalization is due to another, nonphonological factor, such as working memory or Language Background. In this case there would not necessarily be a direct relationship between PM and rule learning.

Experiment 1

Method

Participants

There was a total of 20 participants, 14 undergraduate and postgraduate students and 6 support staff from the University of Cambridge, mean age 29.1 years. There were 11 females and 9 males. None of the participants had formally studied foreign languages beyond the age of 16.

The Language

The participants were required to learn a restricted set of the rules governing determiner selection in Italian. The entire set of items is illustrated in Table 1, where the items reserved for the generalization test are presented in italics. In order to ease the vocabulary learning load, some of the nouns were similar to, or cognates of, English words (the masculine words *tavolo, locale*, and *cliente* and the feminine words *notte* and *stazione*). There were eight determiners, four of which were used with masculine

Table 1

Items employed for training and generalization (in italics) in Experiment 1

	Definite singular (the)	Definite plural (the)	Indefinite singular (a)	Indefinite plural (some)
Paradigm 1. Masculine				
Book	*il libro*	i libri	un libro	dei libri
Present	il regalo	*i regali*	un regalo	dei regali
Bed	il letto	i letti	*un letto*	dei letti
Table	il tavolo	i tavoli	un tavolo	*dei tavoli*
Paradigm 2. Feminine				
Drink	*la bibita*	le bibite	una bibita	delle bibite
Shoe	la scarpa	*le scarpe*	una scarpa	delle scarpe
House	la casa	le case	*una casa*	delle case
Bag	la borsa	le borse	una borsa	*delle borse*
Paradigm 3. Masculine				
Village	*il paese*	i paesi	un paese	dei paesi
Night club	il locale	*i locali*	un locale	dei locali
Saucepan	il tegame	i tegami	*un tegame*	dei tegami
Client	il cliente	i clienti	un cliente	*dei clienti*
Paradigm 4. Feminine				
Paint	*la vernice*	le vernici	una vernice	delle vernici
Night	la notte	*le notti*	una notte	delle notti
Station	la stazione	le stazioni	*una stazione*	delle stazioni
Product	la merce	le merci	una merce	*delle merci*

nouns (*il, i, un, dei*) and four with feminine nouns (*la, le, una, delle*).

Paradigms 1 and 2 in the table are regarded as comprising the regular Italian nouns, where the gender of the noun is indicated by the noun ending (*-o*/*-i* for masculine singular/plural, and *-a*/*-e* for feminine singular/plural). Paradigms 3 and 4 are two of the irregular paradigms in which the noun ending provides no information about the word's gender.

The determiner-noun combinations in italics in Table 1 were reserved for the generalization test and were never presented during training. In the generalization test, participants were required to translate an English phrase (e.g., "a station") into Italian. Given the complex nature of the target system, their choice of determiner could be influenced by both phonological agreement patterns and more abstract knowledge of gender, as is typical of gender systems in natural languages. One highly salient and superficial phonological cue is the euphony between the determiner and noun ending (e.g., *la borsa*). If there were a general "euphony bias" in learning and generalization, then it would differentially affect the various paradigms. But if this were all that participants were sensitive to, then very little influence of the memory factors would be expected, since the fact that determiners and nouns sometimes share the same ending is a very obvious generalization to make from the input data. What is much harder to learn is how this tendency has to be modulated by other item-specific constraints in order to account for the distribution of the determiners. For example, in the case of the singular items in Paradigm 1, there is a unique association between noun inflections and determiners for singulars (specifically, *il*/*un* and *-o*). In Paradigm 2 the determiner-inflection associations are unique if number is also taken into account (*le*/*delle* and *-e*). At a more complex level there is the constraint derived from the attribution of abstract gender to the item, which is clearly most critical in Paradigm 4 to counteract the euphony bias. Both the specific determiner-inflection associations and the abstract gender organization must emerge as generalizations over the training data. Although it would be theoretically interesting to separate the influence of PM on learning the more superficial determiner-inflection associations from its influence on learning the more abstract gender categories, this was not attempted here because of the small number of relevant items (although note that we will focus on the latter in Experiment 2). For present purposes, the point is that whatever the nature of the rules governing generalization test performance, they must be abstracted from determiner-noun combinations encountered during training. The issue is whether those abstractions occur over representations in PLTM.

Procedures

The experiment was conducted using an Apple Macintosh computer. All tasks except the immediate serial-recall task were programmed using Authorware. All sounds were recorded by a female native speaker of Italian. The procedure is summarized in Table 2 and described in detail below.

Language background questionnaire. Before commencing the experiment the participants filled in a questionnaire concerning their foreign language learning background.

PSTM task. The aim of the PSTM task was to test participants' short-term storage capabilities for the vocabulary that would be used in the remainder of the experiment, but at a point where the words were unfamiliar. An immediate serial-recall test was used. The participants were required to immediately recall, in the order of presentation, lists of five of the vocabulary items from the experimental language. The reason for adopting this measure, as opposed to the more standard test of nonword repetition (where a single nonword is immediately repeated), was that it is now well established that PSTM capacity depends upon how similar the test items are to known words (Gathercole & Martin, 1996). What is at issue is the participants' short-term storage capability for the experimental vocabulary (and the relation of that capability to rule learning), and so the

Table 2

Sequence of tasks in Experiment 1

Task	Variable name	Description	Example stimuli → "responses" (all oral unless otherwise stated)
Phonological short-term memory	PSTM	Repeat five-item lists of unfamiliar Italian words in the correct order	libro, tavoli, scarpe, tegame, stazioni → "libro, tavoli, scarpe, tegame, stazioni"
Vocabulary learning	Vocabulary	Recall Italian word given English translation	table → "tavolo," a (printed in red) → "una"
Memory for combinations of morphemes whose forms and meanings are familiar	Input Memory	Recall of "determiner-noun" combinations in sets of three.	context . . . the drinks, le bibite context . . . a book, un libro context . . . some clients, dei clienti RECALL context . . . a book → "un libro" context . . . the drinks → "le bibite" context . . . some clients → "dei clienti"
Rule learning	Generalization	Production of noun phrases given English translation for untrained items, typed responses	context . . . the book, _ libr_ → "il libro"

Note. Input Memory + Generalization test = one cycle. Participants perform three cycles.

only accurate assessment of that ability will be provided by a short-term memory test on precisely those words. Since the vocabulary items were too short for use in a nonword repetition test, the immediate serial-recall method was adopted instead.

The nouns (singular and plural forms) were semirandomly formed into 13 five-item lists with the constraint that no stem form occurred more than once in each list. The nouns were auditorily presented on a computer. Each noun was first played once for the participants for familiarization. Participants were then presented with each list with a pause of 1 s between each word. At the end of each list they were instructed to recall as many words as they could in the order in which they were presented. If they could not remember a word at any position they were instructed to say "pass." There were 3 practice lists, followed by 10 experimental lists.

Vocabulary learning. The participants had to learn to produce the Italian forms of the vocabulary items (nouns and determiners) when provided with English translation equivalents. This test measured the efficiency with which novel phonological forms were established in PLTM. The 16 singular nouns were learned first. On each trial the English translation equivalent of a word was presented on the computer screen. When the participant pressed a key the written and spoken form of the Italian was presented. For the first cycle the participants were instructed to repeat the Italian words. On the following cycles they were told to attempt to recall the Italian when given the English and to then press a key to receive the correct answer. The words were presented in a random order in each cycle. Participants continued cycling through the 16 items until they gave correct responses on all items. Performance on this task provided the basis for the critical measure of vocabulary learning efficiency.

When participants gave correct responses on 100% of the singular nouns, they then proceeded to the plural forms. When they gave correct responses on all of these, they were presented with the singular and plural forms together. When they gave correct responses on 100% of these, they then progressed to the determiners. Because the English determiners had multiple Italian translation equivalents, the different Italian forms were elicited by distinguishing the words for *the*, *a*, and *some* using combinations of different colors and fonts, thus ensuring that each stimulus required a unique Italian response. The assignment of colors and fonts bore no relation to the gender distinction between the determiners. Although somewhat arduous for the participants, this procedure ensured that the forms and meanings of the determiners were effectively learned. Note that only performance on singular nouns was used as the basis of the measure of vocabulary learning efficiency.

Input memory task. In the input memory task, participants were exposed to 48 noun phrases (the nonitalicized items in Table 1). Noun phrases were presented in 16 blocks of three phrases each. Each block consisted of presentation and recall phases. The sequence of events in each block was as follows: An English sentence was presented in which the English translation of the noun phrase was the final constituent. The English noun phrase was printed on a separate line and appeared in boldface (e.g., *Sally liked shopping with her best friend Jane. One Saturday they went shopping together for* **some shoes**). The participant pressed a key and the Italian for the sentence-final noun phrase was visually and auditorily presented (e.g., *delle scarpe*). The participant repeated this phrase. This sequence was repeated for two more noun phrases. The signal "RECALL" then appeared. One of the previous three noun phrases was then presented in English (e.g., *some shoes*) and the participants were required to recall the Italian. They then pressed a key, whereupon they saw and heard the correct answer. This was then repeated for the other two noun phrases in the set. Noun phrases in each set of three were selected so that no determiner or noun appeared more than once. For each participant there was a unique random order of noun phrases in the presentation and recall phases (therefore presentation and recall order were unlikely to be the same). The order of blocks was also individually randomized.

Generalization test. The 16 withheld noun phrases were presented in the generalization test phase. The sequence of events on each trial was as follows: An English sentence was presented in which the English translation of the noun phrase, presented in boldface, was the final constituent (e.g., *The old man was sad that he would never be able to afford* **a house**). Participants attempted to translate the final English noun phrase into Italian by giving an oral response. They were then presented with a display containing the stem of the Italian word and blank fields for the determiner and the inflection (e.g., __*cas*_). They filled in these fields using the keyboard. No feedback was provided.

The input memory task and generalization test constituted one complete cycle. Participants completed three such cycles, yielding three measures of generalization performance at generalization tests 1, 2, and 3. The entire testing session lasted approximately 2 hr.

Results

Overall Scores

Language Background. As anticipated, the participants varied considerably in terms of their knowledge of foreign languages. For this sample the main indicator of foreign language knowledge was the languages studied in secondary school. In the British system, it is usually compulsory to study at least one foreign language (usually French or German) between ages 12 and 16 and to take a General Certificate of Secondary Education (GCSE; formerly O-level) examination in that language. Other languages might also be studied for 1–2 years from age 12, but then the student will either drop them or elect to continue with them to GCSE level. Participants were given two points for any language studied to GCSE level, one point for a language studied for 1–2 years and then dropped, and half a point for a language studied for less than 1 year. The mean score was 2.6 ($SD = 1.5$, range = 1–6).

This dependent variable will be referred to as Language Background. The languages known by the participants (and the number who had at least some knowledge of them) were French (15), German (7), Latin (3), Greek (2), Spanish (1), Hindi (1), and Arabic (1). Note that all but one of the participants had had at least 4 years of instruction and had taken a formal examination in at least one gender language. Because all of the languages known to the participants marked gender distinctions, Language Background scores also reflected knowledge of gender languages.

PSTM. PSTM was scored as follows: Words that were correctly recalled in the correct position were given one point, words in the correct position but with one phoneme incorrect were given half a point, and correct words in the wrong position were also given half a point.[3] Percentages of the maximum score of 50 were calculated. The mean score was 36.5% ($SD = 11.1$, range = 18–55%). The Spearman-Brown split half reliability coefficient was 0.753. This dependent variable will be referred to as PSTM.

Vocabulary. For the Vocabulary measure, responses were scored as correct if there was no more than one phoneme error. The dependent variable Vocabulary was defined as the percentage correct on the third cycle through the singular nouns. (The third cycle was chosen because this was the earliest point at which the fastest learner was almost perfectly correct.) The mean for this variable was 58.9% ($SD = 21.8$, range = 12–94%). In order to gauge the reliability of this measure, the correlation between the scores on the second and third cycles was calculated. The correlation coefficient was 0.879, indicating good reliability.

Input Memory. The dependent variable Input Memory was defined as the percentage of items (determiner-noun pairs) for which the determiner was correctly recalled on the first cycle of the input memory task. Cases in which there were no items between presentation and recall were not counted. Inflection errors were not counted, because they were often difficult to discern (e.g., both final *-e* and final *-a* were often produced as a schwa) and because the correct inflection should have been generated from the English phrase, rather than recalled. The mean for this variable was 74.6% ($SD = 10.8$, range = 54–93%). The Spearman-Brown split half reliability coefficient was 0.678, indicating a rather low degree of reliability.

Generalization. The dependent variable Generalization was defined as the percentage of determiners correctly supplied in the generalization test. Inflection errors were rare (presumably because singular and plural forms had been learned during the vocabulary learning phase) and were ignored. The mean for the first generalization test was 50.0% ($SD = 15.0$); for the second it was 57.5% ($SD = 18.8$); and for the third it was 65.9% ($SD = 19.8$). A repeated-measures analysis of variance (ANOVA) showed that the improvement in performance over cycles was significant, $F(2, 38) = 11.29, p < .001, \eta^2 = .373$. As an indicator of reliability the correlation between the participants' scores at Cycles 2 and 3 was 0.701.[4]

Correlations Between the Memory Measures and Language Background

Correlations between Input Memory, PSTM, and Vocabulary were examined, along with their relationship to Language Background. These correlations are presented in Table 3.

Consistent with previous research, vocabulary learning efficiency was significantly correlated with performance on the

Table 3

Intercorrelations between the memory factors and Language Background in Experiment 1

	Language Background	PSTM	Vocabulary	Input Memory
Language Background	1			
PSTM	.581**	1		
Vocabulary	.38	.51*	1	
Input Memory	.271	.528*	.611**	1

*$p < 0.05$. **$p < 0.01$.

PSTM task. In addition, these two measures both correlated significantly with Input Memory. The participants' ability to remember the determiner-noun combinations was related to their performance on the PSTM task as well as their vocabulary learning efficiency. Language Background correlated significantly with PSTM: The higher the participants' Language Background score, the higher their PSTM.

Correlations With Generalization Test Performance

Table 4 presents the correlations between Language Background, PSTM, Vocabulary, Input Memory, and Generalization at each cycle. Since PSTM correlated with Language Background, partial correlations were computed for PSTM once Language Background had been entered into the regression model, although it should be noted that the accuracy of these partial correlations is compromised by the small sample size.

The predictiveness of the memory measures, particularly PSTM and Vocabulary, was concentrated at Cycle 2, suggesting that they were more strongly related to the rate of learning than its ultimate level. This is to be expected, since if PM determines the quality of the memory representations on which learning is

Table 4

Correlations with generalization test performance in Experiment 1

	Cycle 1 Generalization	Cycle 2 Generalization	Cycle 3 Generalization
Language Background	.552*	.679**	.640**
PSTM	.458* (.202)	.601** (.346)	.340 (.052)
Vocabulary	.330	.633**	.435
Input Memory	.188	.444*	.493*

Note. Partial correlations after Language Background has been entered are shown in parentheses.
*$p < 0.05$. **$p < 0.01$.

based, then learners with poor PM will tend to require more exposure than those with good PM in order to achieve the same level of learning. As the amount of training increases, so the amount of variance in learning scores accounted for by memory factors will be bound to decrease.

Input Memory was correlated with Generalization at Cycles 2 and 3, although not very strongly. We consider the possible reasons for this below. The partial correlations involving PSTM were not significant at any point. It is therefore not clear whether there was an underlying effect of PSTM that was being masked by its intercorrelation with Language Background or whether the simple correlations between PSTM and Generalization were really due to Language Background.

The strength and consistency of the correlations between Language Background and Generalization suggest that non-memory factors that were reflected in this measure were a major determinant of rule learning. Indeed, Language Background was a stronger determinant of the ultimate level of learning achieved at Cycle 3 than any of the memory measures.

Discussion

As predicted on the basis of previous research, PSTM correlated significantly with vocabulary learning efficiency, demonstrating that the ability to recall unfamiliar words in the short term is related to the ability to form phonological representations that persist over the long term. The fact that we found this relationship suggests that the present immediate serial-recall task is as valid a measure of PSTM as the nonword repetition task used by other researchers (Gathercole et al., 1992; Gathercole, Willis, Baddeley, & Emslie, 1994; Service & Craik, 1993). We shall treat both PSTM and Vocabulary as measures of PM ability; that is, the ability to store novel phonological forms in PSTM and establish corresponding representations in PLTM.

Both of the above measures of PM were correlated with Input Memory. It will be recalled that previous research on memory for familiar forms has suggested that beyond the approximate 2 s duration of PSTM, longer term storage is in a non-PM system. But these experiments were performed using stimuli with unique meanings. This made it possible for participants to remember the input sequences using semantic representations. In the present case the problem was not remembering the meaning of the words that occurred, because these were provided as recall cues in the form of the English translations. The problem was remembering the specific phonological form of the word that was used to express the meaning, for example, whether the word for "the" in "the station" was *il*, *i*, *la*, or *le*. The participants therefore had to rely on PM. Since the recall interval was always greater than the assumed duration of PSTM, we assume that recall was supported by the representations of the specific determiner-noun combinations that were established in PLTM.

Although we use the term "phonological long-term memory" to refer to this memory system, this is not meant to imply that recall would be immune to delay in the sense normally associated with the concept of long-term memory. The present recall task required retention of the input combinations over at most four intervening events (when the first presented item was the last to be recalled). There was no measure of the participants' recall capability over longer intervals. However, we will assume that in the early stages of learning, after the first few exposures to an item, the PLTM representations that are formed are subject to decay or interference in much the same way as they are in PSTM. This would limit the interval over which recall would be possible. With more repetitions of the item, the PLTM representation would be strengthened until the point at which recall is possible at any delay. Of course, this raises the issue of whether there is in fact a distinction between PSTM and PLTM at all and whether there is just a gradient of representation strength within a unitary PM system. This is currently an unresolved issue in the memory literature: Compare for example the memory models of Cowan (1993) and Baddeley (1990). For present purposes we merely assume that the strength of encoding in PLTM was sufficient to support rule learning.

The measure of input memory correlated significantly with rule learning as measured by the generalization tests. However, the relationship was not as strong as one might have expected on the assumption that it is memory for input that is the driving force in learning. A relevant consideration here is the rather low reliability of the Input Memory measure. This may reflect the potential for variation in the factors that determined recall from trial to trial. Because recall was cued with the English translation, it was possible for participants to reduce the number of response alternatives by bearing in mind the meaning of the required determiner. If they were cued with "the," then their choice of determiners was reduced to four (*il, i, la, le*), whereas if cued with "a" or "some," their choice was reduced to two (*un* or *una* and *dei* or *delle*, respectively). This increased the chance probability of a correct response even in the absence of any recollection of the actual input item. Moreover, it did so in a way that varied from trial to trial according to the item and also potentially according to whether the participant paid attention to semantic cues at recall. This could have contributed to the unreliability of this measure. In Experiment 2 an alternative test of memory for input was devised that was intended to be more sensitive to participants' actual memory for the input items.

The correlations that were obtained between Input Memory and Generalization do not by themselves suggest that rules were abstracted from representations of input in PLTM. Clearly Input Memory has a PM component, as shown by its relationship to PSTM and Vocabulary (Table 3). But it also needs to be shown that there is a direct relationship between measures of PM (PSTM/Vocabulary) and generalization test performance for it to be proven that it is the PM component of Input Memory that relates to rule learning. There was some evidence for this in the form of a significant correlation between Vocabulary and Generalization at Cycle 2. However, no significant partial correlations were obtained for PSTM.

One unexpected feature of the results was the strong influence of Language Background, here operationalized simply as the extent of the participants' knowledge of foreign languages, all of which involved gender systems. It is striking that it is the least complex of the memory tasks, the PSTM task, which was influenced by this factor (see Table 3), making it difficult to assess its precise contribution to rule learning. One reason for the influence of language background on PSTM may have been that participants who knew many languages may have been more likely to exploit similarities between the vocabulary items used in the experiment and words in other languages that they knew. It is known that performance on PSTM tasks with unfamiliar words is influenced by the extent of vocabulary knowledge in both native and nonnative languages (Thorn & Gathercole, 1999).

Language Background was also strongly related to Generalization. It is not unlikely that the more gender languages a participant knew, the more he would be aware that gender is an important organizing principle, particularly in Romance languages. There might also have been a direct influence from knowledge of French, which shares the determiners *la* and *le* with the Italian system (although only the former has the same function). In Experiment 2 we sought to reduce the influence of language background on both PSTM and rule learning by employing nonsense words with a lower level of similarity to languages that the participants were likely to know.

Experiment 2

In Experiment 2 a number of procedural and design changes were made in order to refine the measures of PM and to reduce the influence of participants' language background. The main change was with respect to the language employed in the experiment. Experiment 2 utilized an invented language in which the nouns and determiners were nonwords chosen so as not to be similar to words in languages that the participants were likely to know. The new target language was an analogue of Paradigms 3 and 4 from the language used in Experiment 1. Noun endings now provided no cue to word class. As noted previously, generalization test performance in Experiment 1 could have been based on a combination of item-specific determiner-inflection associations and knowledge of abstract gender class, as well as a general euphony bias. Determiner-inflection associations are relatively simple because they depend upon first-order associations between morphemes that are encountered directly in the input items. Knowledge of abstract gender is more complex because it depends upon second-order associations. Nouns become grouped together in classes by virtue of their associations to the same set of determiners, and determiners become grouped by virtue of their associations to the same set of nouns (Maratsos, 1982). This involves forming associations between items that never occurred directly together in the input. It has been argued that distributional analysis of this kind is a powerful method that language learners utilize to break into lexical as well as grammatical structure (Redington & Chater, 1998). Here we wanted to test specifically whether PM relates to this more complex form of distributional learning.

In addition a number of changes were made in order to sharpen the memory measures:

> *PSTM.* In the immediate serial-recall task of Experiment 1, an auditory-only presentation method was used, whereas in the vocabulary learning and input memory tasks, the target language was presented both visually and auditorily. Thus, performance might have been affected by low-level problems of phoneme identification in a way that performance in the other tasks was not. Therefore, in the immediate serial-recall task of Experiment 2, items were presented both visually and auditorily. The number of items that had to be recalled was also reduced from five to three. In long lists, accuracy of serial recall is a function both of the ability to retain the items in the list and also of the ability to retain their serial position. It is possible that these are two separate

abilities, yet for present purposes it is the ability to retain the phonological form of the items that is of most interest. The length of the lists was therefore reduced to three items in order to ease the problem of remembering serial order. Finally, as mentioned above, the use of nonwords with little resemblance to words in languages that the participants were likely to know was intended to reduce the impact of language background on PSTM scores, enabling a more straightforward assessment of the relationship between PSTM and other factors.

Input Memory. In this experiment a measure of input memory was taken prior to the vocabulary learning phase. The participants were first familiarized with the forms of the morphemes, but *not* their meanings, by having them complete a fragment completion task. The participants then performed an intermediate-term memory task on "determiner-noun" combinations (which were not recognizable as such at this point). This provided a measure of memory for morpheme combinations that could not be affected by translation strategies. Recall was possible only by relying exclusively on memory for the items themselves. This new memory measure will be referred to as Memory for Morpheme Combinations. Clearly this change to the procedure changed the nature of the vocabulary learning phase, which merely required familiar forms to be associated with meanings.

Method

Participants

There were 21 participants, 19 students and 2 support staff from the University of Cambridge, mean age 23.7 years. There were 10 females and 11 males. None of them had formally studied foreign languages beyond the age of 16.

The Language

An analogue of Paradigms 3 and 4 from Experiment 1 was employed. The training and generalization items (the latter in italics) are shown in Table 5.

A Japanese dictionary was used to provide the basis for many of the items (which were then modified). All singular forms ended in -*e* and all plural forms ended in -*i*. All inflected nouns contained three syllables. The determiners were based on the Italian determiners, but the consonants were systematically changed: the letter *l* was changed to *g*, *n* was changed to *l*, and *d* was changed to *t*. Thus, for example, *tegge* was derived from *delle*, and *ul* was derived from *un*. As in Experiment 1, each word occurred with only three of the possible determiners during the memory task, the fourth being withheld for the generalization tests.

Procedures

All of the tasks were programmed in Authorware on a Macintosh computer. Sounds were recorded by the first author

Table 5

Items employed for training and generalization (in italics) in Experiment 2

	Definite singular (the)	Definite plural (the)	Indefinite singular (a)	Indefinite plural (some)
"Masculine"				
Ball	*ig johombe*	i johombi	ul johombe	tei johombi
House	ig zabide	*i zabidi*	ul zabide	tei zabidi
Fight	ig wakime	i wakimi	*ul wakime*	tei wakimi
Bird	ig migene	i migeni	ul migene	*tei migeni*
"Feminine"				
Shoe	*ga shosane*	ge shosani	ula shosane	tegge shosani
Kiss	ga tisseke	*ge tisseki*	ula tisseke	tegge tisseki
Cake	ga chakume	ge chakumi	*ula chakume*	tegge chakumi
Nose	ga nawase	ge nawasi	ula nawase	*tegge nawasi*

(a native speaker of British English). Before commencing the experiment, the participants filled in a questionnaire concerning their foreign language learning background. The sequence of tasks is summarized in Table 6.

PSTM. There were actually three immediate serial-recall tasks, administered in the order below, but only those involving singular and plural nouns constituted the critical PSTM measure, whereas that involving determiners served for item and task familiarization.

> *Determiners.* The eight determiners were semirandomly formed into 13 three-item lists. No determiner could occur more than once in each list, and the distance between repetitions of the same determiner across lists was kept roughly constant. In contrast to Experiment 1, the items were presented both auditorily and visually with a delay of 1 s between the acoustic offset of one item and the onset of the next. After the third item the word "RECALL" flashed on the screen, and the participant attempted to recall the items in the correct order.
>
> *Nouns.* As above, but the singular forms of the nouns were used. This was then followed by 13 more lists containing plural nouns. Performance on these tasks constituted the PSTM measure.

Fragment completion task. The fragment completion task was designed to teach the participants the forms of the words without introducing their meanings. Participants were provided with the consonants corresponding to the first phoneme of the words, and they had to recall the whole item. In order to disambiguate the determiners, blanks corresponding to the number of letters were also provided. For nouns, a final *-e* or *-i* was also provided after a continuous line representing the stem. This was to give the participants practice at retrieving both forms of the nouns. Participants practiced this task using a mixture of computerized presentation and flash cards until they achieved 100% accuracy on the computerized version.

Table 6

Sequence of tasks used in Experiment 2

Task	Variable name	Description	Example stimuli → "responses"
Phonological short-term memory	PSTM	Repeat three-item lists of unfamiliar words in the correct order	ul, ig, ga → "ul, ig, ga" chakume, johombe, nawase → "chakume, johombe, nawase"
Fragment completion		Recall whole word given initial/final letters	_l → "ul" c___e → "chakume"
Memory for combinations of morphemes whose forms, but not meanings, are familiar	Memory for Morpheme Combinations	Recall of "determiner-noun" combinations in sets of three	ul johombe, ge shosani, ig migene _ shosan_ → "ge shosani" _ johomb_ → "ul johombe" _ migen_ → "ig migene"
Vocabulary learning	Vocabulary	Recall word given English translation	bird → "migene" a → "ul, ula"

Table 6 (continued)

Sequence of tasks used in Experiment 2

Task	Variable name	Description	Example stimuli → "responses"
Memory for combinations of morphemes whose forms and meanings are familiar	Input Memory	As for Memory for Morpheme Combinations, but with four-item sets and English translation as retrieval cue	the nose, ga nawase the birds, i migeni some balls, tei johombi a kiss, ula tisseke the birds, _ migen_ → "i migeni" the nose, _ nawas_ → "ga nawase" a kiss, _ tissek_ → "ula tisseke" some balls, _ johomb_ → "tei johombi"
Rule learning	Generalization	Production of noun phrases given English translation for untrained items, typed responses	the ball, _ johomb_ → "ig johombe"

Note. Input Memory + Generalization test = one cycle. Participants perform a maximum of five cycles.

Memory for morpheme combinations. Determiner-noun combinations from the training set shown in Table 5 were presented in blocks of three. Each item was presented both auditorily and visually. At sound offset, the visual form was replaced by question marks, and the participants were required to repeat the noun phrase aloud. They then pressed a key to receive the next item. After the third item the word "RECALL" appeared, the participants pressed a key, and the stem of one of the items was visually presented, flanked by underscores at the positions of the determiner and inflection (length of the determiner underscore was constant for all items). They attempted to recall the whole item aloud, then pressed a key to see and hear the correct answer. This feedback was followed by the stem for the next noun phrase to be recalled. The assignment of noun phrases to blocks and the randomization of presentation and recall orders were unique for each participant and each cycle through the training set. The program assigned noun phrases to blocks following the constraint that each stem and each determiner should appear only once in each block. Presentation order was random, and recall order was semirandom, the constraint being that the last noun phrase to be presented should not be the first to be recalled. Participants performed two cycles through the 24-item training set in this task.

Vocabulary learning. The meanings of the eight singular forms of the nouns were learned in the same way as in Experiment 1. The plural forms were learned using flash cards until the participant could produce the translation for each English word. The determiners were learned in the same way. Each card had an English determiner on one side and all of the translation equivalents on the other. For the word "the," there were four cards with the order of the translation equivalents different on each one, and for the words "a" and "some," there were two cards with two different orderings of the translations. The participants were required to remember the translation equivalents, but not their order. Different orders were merely used so as not to give any particular ordering significance.

Input memory task. The input memory task followed the same procedure as the morpheme combination task, except that items were now presented in blocks of four rather than three (because pilot work showed that with the addition of meaning cues, three-item blocks were too easy). The English translation appeared 1 s before each noun phrase (to speed up the procedure, sentence contexts were not used in this experiment). The English translation remained on the screen during presentation and immediate recall of the noun phrase and was presented again along with the stem at recall. The same randomization procedure was used as in the morpheme combination task.

Generalization test. In the generalization test, the withheld items from Table 5 were presented in random order. The procedure was the same as in Experiment 1, except no sentence contexts were used.

The sequence of input memory task and generalization test comprised one cycle. Participants completed five such cycles, yielding five measures of generalization performance (the amount of exposure was increased with respect to Experiment 1 because the present system was predicted to be more difficult to learn). Participants were permitted to finish prior to completing five cycles if either (a) they had achieved 100% on two consecutive cycles or (b) their score was still at chance (37.5%) on the fourth cycle and they had not exceeded this score on any previous cycle.[5] For the purposes of statistical analysis, in cases in which participants were permitted to finish prior to completing five cycles, the scores for the remaining cycles were entered as the same as that on the last completed cycle. The entire testing session lasted approximately 2 hr.

Results

Overall Scores

Language Background. As in Experiment 1, the participants varied considerably in terms of their knowledge of foreign languages. The same system for rating this knowledge was used as in Experiment 1. The mean score was 3.05 ($SD = 1.5$, range $= 0$–6). All but one of the participants had some knowledge of a gender language, and all but two of them had had at least 4 years of instruction and had taken a formal examination in at least one gender language. The languages known by the participants (and the number who had at least some knowledge of them) were French (16), German (7), Latin (4), Spanish (3), Greek (1), Korean (1), and Shona (1). Apart from Korean and Shona, all of the languages known to the participants were gender languages.

PSTM. PSTM was scored as follows: Words that were correctly recalled in the correct position were given one point, words in the correct position but with one phoneme incorrect were given half a point, and correct words in the wrong position were given half a point (such errors were rare). The mean over the two immediate serial recall tasks using nouns (singulars and plurals) was 64.5% ($SD = 17.8$, range $= 31$–95%). The Spearman-Brown split half reliability coefficient was 0.939, indicating that this was a highly reliable measure.

Memory for Morpheme Combinations. Memory for Morpheme Combinations was defined as the percentage of items (determiner-noun pairs) for which the determiner was correctly recalled. The mean score was 45% ($SD = 11.7$, range $= 29$–71%). The Spearman-Brown split half reliability coefficient was 0.809, indicating a high degree of reliability.

Vocabulary. Because learning was extremely rapid, the dependent variable Vocabulary was defined as the percentage of nouns recalled on the first recall cycle, rather than the second. The mean was 79% (SD $= 20$, range $= 25$–100%), and the average number of retrieval cycles required to reach 100% correct was 2.1 ($SD = 0.79$, range $= 1$–3). Clearly, the familiarity of the forms rendered the task of mapping them onto meanings almost trivial.

Input Memory. Input Memory was defined as the percentage of items for which the determiner was correctly recalled in the input memory task on the first cycle. The mean score was 69% ($SD = 10.9\%$, range $= 42$–87). However, the Spearman-Brown split half reliability coefficient was only 0.11, indicating an unacceptably low level of reliability. This measure was therefore dropped from all subsequent analyses.[6]

Generalization. The percentage of determiners correctly supplied in the generalization test after each cycle of the learning phase constituted the measure of rule learning. The mean percentages correct over the five tests were 36%, 48%, 54%, 66%, and 67%. A repeated-measures ANOVA showed that the improvement in performance was significant, $F(4, 80) = 13.11, p < .001, \eta^2 = .396$. As an indication of reliability, the correlation between the scores on the penultimate and final cycles was 0.803.

Intercorrelations Between the Memory Measures and Language Background

The correlations between Language Background, PSTM, Vocabulary, and Memory for Morpheme Combinations were calculated. The only significant correlation was between Memory for Morpheme Combinations and PSTM, $r = .567, p < .01$. There were no correlations involving Language Background; in particular, the correlation between PSTM and Language Background was 0.126. The correlation between PSTM and Vocabulary was 0.104, suggesting that this correlation was no longer sensitive to PM.

Correlations With Generalization Test Performance

Correlations with Generalization at all five cycles are shown in Table 7. Vocabulary was not related to ultimate learning, which is not surprising given that it no longer appeared to be related to PM. Both PSTM and Memory for Morpheme Combinations predicted Generalization in the later cycles. Crucially, PSTM was correlated with Generalization on the final cycle. At Cycle 5 the partial correlation between PSTM and generalization test performance once Language Background had been removed was 0.545, $p < .05$, although again the accuracy of this estimate is compromised by the small sample size. The strongest predictor of ultimate learning was Memory for Morpheme

Table 7

Correlations with generalization test performance in Experiment 2

	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
LANG	.450*	.287	.287	.593**	.520*
PSTM	.006	.359	.369	.344	.528*
Vocabulary	.215	.091	.527*	.308	.290
Memory for Morpheme Combinations	.036	.111	.112	.468*	.641**

*$p < 0.05$. **$p < 0.01$.

Combinations, the new, more reliable measure of the ability to remember morpheme combinations that was introduced in this experiment. As in Experiment 1, Language Background was also correlated with Generalization, despite the fact that on this occasion it was not correlated with any of the memory measures. In other words, there was an effect of Language Background on learning that was not mediated by memory factors.

Discussion

The Memory for Morpheme Combinations task was introduced in order to measure the ability to store novel combinations of familiar morphemes without contamination from translation strategies. This new measure was found to be correlated with PSTM, suggesting that the ability to store novel combinations of familiar morphemes in PLTM is related to PSTM. The significant correlations between Memory for Morpheme Combinations and Generalization at Cycles 4 and 5 suggest that the ability to store morpheme combinations relates to ultimate rule learning. These results are similar to those obtained in Experiment 1. In the present experiment there was also a direct correlation between PSTM and Generalization at Cycle 5. The pattern of results suggests that Memory for Morpheme Combinations relates to rule learning because it reflects the participants' ability to encode input sequences reliably in PLTM.

The change in the language used reduced the impact of Language Background on the memory measures, especially PSTM. This eliminated the confound between these two factors, and the relationship between PSTM and generalization test performance became apparent. Even though Language Background continued to predict learning outcomes, an independent effect of PSTM was obtained. The reduction in the correlation between Language Background and PSTM is presumably due to the use of nonwords, which were less likely to be similar to other words that the participants knew than the Italian words used in Experiment 1. Nevertheless the relationship between Language Background and rule learning persisted. This suggests that it was some nonmemory component of Language Background that was related to rule learning. The possible nature of this nonmemory component will be discussed below.

The results of Experiment 2 contrast with those of previous studies that have found that arbitrary word classes are not learnable (Braine et al., 1990; Brooks, Braine, Catalano, & Brody, 1993; Frigo & McDonald, 1998). On the other hand, Brooks et al. (1993) and Frigo and McDonald (1998) found that participants could learn a noun class distinction when 60% of the nouns carried a distinctive phonological cue to their class. Frigo and McDonald reject as being "too powerful" (p. 237) models of noun class learning that depend on pure distributional analysis (Anderson, 1983; Maratsos & Chalkley, 1980; Pinker, 1984). But here we have found that an arbitrary system is learnable by at least some participants. Learning by pure distributional analysis is possible, but it requires a good memory. Certain models of learning may be too powerful in the sense that they assume perfect memory for input, but this does not invalidate the proposed learning mechanisms.

It is possible that the discrepancy with previous studies is due to the size of the languages involved. Braine et al. (1990) used a 24-word vocabulary, Brooks et al. (1993) used 30 words, and Frigo and McDonald (1998) used 20 words, whereas Experiment 2 used only 8 words. Clearly, keeping track of the collocates of 20 to 30 words is much harder than keeping track of the collocates of 8 words. Another factor that may have facilitated learning in Experiment 2 is that the determiners themselves bore weak similarity relations to each other (e.g. *ga–ula*, *ge–tegge*, *i–tei*). Such similarities may have made it easier to form associations between the determiners than in the artificial languages used in the previous studies.

General Discussion

The present experiments found correlations between PM ability and rule learning. This is consistent with the assumption that PM abilities are related to language learning (Baddeley et al., 1998; Ellis, 1996). However, as noted in this article's introductory section, prior studies have not separated the effect of PM on morpheme learning from its effect on rule learning. The present experiments showed effects of PM on rule learning even when all of the relevant morphemes were familiar. Furthermore, Experiment 2 demonstrated that these effects could be obtained even when generalization test performance depended on an abstract categorization of nouns into word classes, as opposed to direct associations between morphemes that occurred together in training items. Thus, PM appears to play a role in learning even quite abstract aspects of grammar.

Given that PSTM was shown to be correlated with measures of memory for input (Input Memory in Experiment 1 and Memory for Morpheme Combinations in Experiment 2) and that memory for input was shown to be related to rule learning, it seems reasonable to conclude that PM influences rule learning via its influence on input memory. Previous research has established the relationship between PSTM and the ability to store the phonological forms of novel words in PLTM. The present data suggest that PSTM also relates to the ability to store novel combinations of familiar morphemes in PLTM and to use these as the basis for rule learning.

The ability to form associations between familiar phonological forms is clearly critical to learning in this domain, at the level both of first-order associations between noun endings and determiners and of second-order associations among the nouns and determiners themselves. But this emphasis on associative learning should not be taken to imply that the resulting generalizations automatically and implicitly emerge from memory for input. In fact, we have good reason to believe that the learning processes occurring in these experiments were largely explicit. Informal debriefings conducted after Experiment 2 showed that the 6 participants who scored 100% on the final generalization test were able to describe clearly the organization of the determiners and the separation of the nouns into two classes. In contrast, none of the 12 participants who scored 62.5% or less in the final generalization test were able to group all of the determiners into the correct classes, although the 3 participants who scored 62.5% did understand the singular-plural opposition between *ig* and *i* on the one hand and *ga* and *ge* on the other, showing that they had partial knowledge of the system. There was thus a good correspondence between participants' own conscious understanding of the rule system and their ultimate generalization test performance.

Another reason for believing that, in Experiment 2 at least, learning was dependent on explicit processes of some kind is that in subsequent experiments (Williams, in press), it was found that the same language was not learnable under implicit/incidental task conditions. In those experiments, there was no recall task; participants were required to make a semantic decision about each phrase and then to translate it into English. They were encouraged to perform these tasks as quickly as possible. The participants had considerably more knowledge of gender languages than those in the present experiments (mean Language Background score = 5.8, compared to 2.6 in Experiment 1 and 3.0 in Experiment 2 of the current study), and they were provided with three times more exposure (15 cycles through the training set). Yet performance on the generalization test was at chance. Crucially, this was the case even for participants who had very good memory for trained items.[7] The trained items had been encoded in long-term memory, but this information by itself was not sufficient to support learning of the noun class distinction. Of course, it may be that learning under implicit conditions is so slow that it was not observed within the time scale of the experiment, or else higher levels of memory for trained items might be required for generalizations to emerge implicitly. Nevertheless, the fact that in the present Experiment 2 learning was obtained after less exposure suggests that explicit learning processes were important.

This emphasis on explicit learning might seem surprising given that we used training tasks that appeared to be exercises in rote memorization. Even the generalization tests could also have been perceived as tests of long-term rote memory, because participants did not appear to notice that the items had never been encountered during training. However, these tasks can equally be regarded as pushed output and reconstruction tasks of the type that have been shown to stimulate explicit learning processes. For example, Izumi, Bigelow, Fujiwara, and Fearnow (1999) showed that opportunities for repeated reconstruction had positive effects on noticing task-relevant aspects of linguistic form, and there was some evidence of enhanced rule learning in relation to those forms. In addition, because after each recall attempt our participants were provided with the correct response, the present task can be regarded as incorporating a form of interactional feedback, similar to recasts. Mackey (1999) showed that direct involvement in tasks involving interactional feedback had beneficial effects on language development. According to the output hypothesis (Swain & Lapkin, 1995), pushed output is beneficial because it forces the learner to think about aspects of form. According to the interaction hypothesis (Long, 1996), interactional feedback is beneficial because it stimulates noticing of form. Recasts are assumed to be beneficial because they allow learners to notice, and analyze the reasons for, the gap between their own production and the target form (Gass, 1997).

How might explicit learning processes be involved in learning the present rule systems? Here we can only speculate, because the nature of the rules that are learnable only by explicit mechanisms has yet to be adequately defined (although for evidence of the failure to learn abstract, nonlinguistic rules implicitly, see Mathews et al., 1989; Perruchet, 1994; Shanks, Johnstone, & Staggs, 1997). However, as explained earlier, a crucial aspect of learning arbitrary noun class distinctions is that second-order associations need to be formed between elements that never occur together in the input. Nouns of the same class become associated because they occur with the same determiners, but the relevant training items are encountered at different times. Likewise, determiners can become associated because they occur with the same nouns, but there may be many learning trials between the relevant events (indeed the randomization procedures used in Experiments 1 and 2 ensured that items containing the same determiner or noun did not occur in the same block of trials). It is possible that explicit comparison processes are required for the mediated associations to be acquired, at least within the time scale of these experiments.

But does an appeal to explicit learning processes reduce the relevance of PM? We would argue that it does not. Explicit comparison processes obviously require that participants have good memory for determiner-noun combinations encountered during training, and as we have shown, this is related to PM ability. The new associations that are formed as a result of explicit comparison also concern novel combinations of (familiar) phonological forms, and so storage of these should also be dependent upon PM.

An appeal to explicit learning processes allows us to interpret a salient aspect of the present results: Language learning background exerted a clear and consistent influence on rule learning that was independent of memory abilities. A similar influence of language background on artificial grammar learning was reported by Nayak, Hansen, Krueger, and McLaughlin (1990), who tested the connection between multilingualism and language learning in a laboratory experiment using an artificial, but natural-like, phrase structure grammar comprising strings of nonsense words (e.g., *rud tiz jax neb*, *mik tiz vot*) paired with geometric figures. When rule discovery instructions were provided, the multilinguals outperformed the monolinguals. This advantage was unlikely to be due simply to the multilinguals' greater propensity to search for rules, because participants' verbal reports showed the same level of use of rule discovery strategies in the two groups. Nor was it due to superior memory abilities, because the two groups did not differ in vocabulary learning (the extent to which they picked up the associations between the nonwords and geometric figures). This suggests that the two groups were matched in terms of PM ability. Nayak et al. (1990) conclude that the multilingual advantage in rule learning was "most likely a reflection of their experience with language and their willingness and ability to search for rules" (p. 241). Language background may reflect a general "willingness to analyze" the input and is perhaps related to language learning aptitude (Skehan, 1998). However, in the present context a more specifically relevant aspect of "experience with language" could be related to knowledge rather than aptitude. Knowledge of other gender systems could provide an abstract understanding of the way that nouns can be grouped into classes that are defined purely in terms of their distributional properties. This could be a crucial factor in determining whether participants adopt the kind of explicit analysis strategy outlined above, that is, one that involves making comparisons between disparate input items in order to form second-order associations.

We conclude that in the present experiments both memory and nonmemory factors affected learning under task conditions in which explicit processes were having some effect. These results are therefore similar to those obtained by Robinson (1997) in his instructed and rule search conditions, in which rule learning was jointly determined by memory and nonmemory (and in that case "grammatical sensitivity") factors. Both sets of results illustrate how rule learning is not determined solely by memory ability. It cannot be assumed that rules emerge from memory representations of input sequences in quite as straightforward a way as implied by emergentist approaches to language learning, at least not for the kinds of learners and under the kinds of conditions in these experiments.

A similar conclusion was reached by Williams (1999). In that study, as here, a measure of PM (vocabulary learning efficiency) was related to rule learning. However, the degree of learning in a specific domain, for example, article-noun agreement, was unrelated to recall accuracy for the relevant forms early in the experiment. In the present terminology, PM was related to learning, but input memory was not. It was suggested that this was because of a high degree of variability in encoding of input sentences before participants had developed suitable strategies for integrating the various elements of the sentences in a way that was conducive to learning. In the present experiments, the problem of encoding individual input sequences was lessened because of the narrow focus on determiner-noun pairs, although note that input memory was still a relatively unreliable memory measure. But the problem of achieving an appropriate integration between input elements for learning was if anything made more acute because, as pointed out above, the relevant elements were dispersed over the entire set of training items. In both cases memory ability is clearly important, but so too is the ability to relate different aspects of the input to one another in a way that is relevant to the learning problem.

Do these results have more specific implications for learning gender systems in natural languages? There is a growing body of evidence that suggests that even quite advanced L2 learners continue to make gender errors (Holmes & de la Batie, 1999). In contrast, such errors are relatively rare in L1 acquisition (Caselli, Leonard, Volterra, & Campagnoli, 1993). There is also evidence for qualitative differences between L1 and L2 acquisition and processing. A number of studies have shown that L2 learners are more sensitive to phonological agreement patterns that correlate with gender classes than either children or adults in their native language. For example, for the Italian *il pettine* 'the comb,' an L2 learner might produce **le pettine*

(Holmes & de la Batie, 1999). Such an error suggests that determiner selection is influenced by the noun ending. In contrast a child would be more likely to produce **il pettino*, choosing an article that is correct for the noun's gender and providing the noun with the characteristic ending for that class. This demonstrates a grasp of the noun's abstract gender as the controlling influence in determiner selection (Caselli et al., 1993). In reaction time tasks on adults, Taraban and Kempe (1999) showed that nonnative speakers of Russian are more sensitive to phonological cues to gender than are natives. Finally, a gender-priming study by Guillelmon and Grosjean (2001) showed that whereas native speakers of French and early bilinguals show gender congruency effects in reaction time tasks, such effects are absent in late bilinguals. This too suggests that L2 learners do not achieve native-like representation or processing of gender information.

Many of the present participants clearly had difficulty inducing gender classes, and it has been argued that both memory and nonmemory (conceptual) factors were responsible. We have suggested that the conceptual factor relates to the participants' ability to understand that gender classes can be defined by purely distributional criteria. This problem would presumably influence real-life language learning in primarily inductive, rather than instructed, contexts. Even at the earliest stages of formal instruction in a language, however, the learner should be aware that certain determiners are indicators of gender. Yet problems with gender are reported even for individuals with a high level of linguistic sophistication (Carroll, 2001, p. 358). In such cases the problem simply seems to be in remembering which determiners occur with which nouns. This points to problems in storing determiner-noun combinations in PLTM.

But if adult L2 learners' difficulty inducing gender and learning the gender of specific nouns is due to these conceptual and PM factors, why are similar problems not evident in L1 acquisition? It is evidently not the case that children have the kind of metalinguistic understanding of noun class systems and their distributional properties that underlies the inductive learning strategies of our more successful participants. Nor is it the case that children have superior PM ability, because performance on PSTM tasks, such as nonword repetition, improves between the ages of 4 and 8 (Gathercole et al., 1994). It is precisely when PM ability is at its most limited that problems with gender assignment are least evident. This paradox might suggest that a grammatical concept of gender is fully available only in L1 acquisition (Carroll, 1989). In inducing a noun class distinction, the child might benefit from a propensity to interpret specifiers as potential indicators of noun class membership, that is, as assigners of gender features. Once the child has realized that particular forms, such as determiners, are indicators of gender class, these can be used to assign an abstract gender feature to the nouns with which they occur in the input (Carroll, 2001, pp. 357–362). The gender feature would be attached to the representation of the noun at the lemma level (Levelt, Roelofs, & Meyer, 1999), providing a robust encoding of its grammatical class. Since lemmas are assumed to be abstract, nonphonological representations, storage at this level would not be constrained by PM ability. In contrast, if the L2 learner no longer has access to the gender feature, then he must construct gender classes from the network of first- and second-order associations between morphemes that is derived from representations of input sequences in memory. The present results suggest that these representations are stored in PLTM, and so variation in PM ability might underlie differences in the ability to remember the gender of L2 words.

There is thus an interesting paradox when L1 and L2 acquisition of gender is considered in the light of child-adult differences in PM and metalinguistic knowledge. As was noted in the present article's introductory section, it has been argued that PM does not relate directly to the acquisition of grammatical knowledge in the L1, although it might relate to spoken language ability (Adams & Willis, 2001). The account of L1 acquisition of gender outlined above fits in well with this more general hypothesis. At the same time our results fit in with an emerging consensus that PM is related to adult L2 grammar learning (Ellis et al., 1999; Ellis & Sinclair, 1996; Robinson, 1997; Williams, 1999). However, although the notion of the gender feature is appealing, other approaches to explaining the difference between L1 and L2 acquisition of gender should also be pursued. For example, differences in sensitivity to phonological cues may be accounted for simply in terms of degree of exposure and asymptotic learning (Taraban & Kempe, 1999). Or the L2 learner's problem may be a consequence of the way in which L2 vocabulary items are stored in relation to the deeply entrenched lexical representations and organizational structures developed in the course of L1 acquisition, an issue that may be fruitfully explored using connectionist models.

The present research represents only a small step toward understanding the role of memory processes in rule learning. One issue for future research is the role of PM as opposed to more complex working-memory processes of the type represented in the work of Just and Carpenter (1992) and Engle et al. (1999). Given that the training tasks used here are similar in some respects to working-memory tasks, and given that learning appeared to depend upon explicit processes, it is possible that part of the predictiveness of the input memory measures derived from a working-memory, rather than a PM, component. Future research would ideally include tests of both PM and working memory, and with larger sample sizes than were used here it would be possible to explore the relationships between the various predictors using hierarchical regression or factor analysis.

Another issue is that of the nature of the rules. Here we examined the role of PM in learning rules that obviously depend on associations between phonological forms. But what of rules that concern form-meaning, as opposed to form-form, relationships, as for example in the kind of semantically defined noun class system found in languages such as Tamil or in the case of the complex semantic conditions governing English article usage? And what of other rules that depend upon the distribution of grammatical classes rather than phonological forms as such? Here we concentrated on a system in which phonological form had to be retained for the relevant rules to be learned. But this is not strictly the case for many syntactic rules. Clearly it is necessary to examine rule systems of this type in order to arrive at a more complete picture of the role of PM in grammar learning.

Revised version accepted 21 June 2002

Notes

[1]For example, if a person wants to remember a 10-digit telephone number long enough to dial it, he may continually repeat or rehearse it to himself. But if by the time the 10th digit has been rehearsed the first digit has been lost from memory, due either to decay or to interference, then recall of the entire sequence will not be possible. Thus, even when a rehearsal strategy is used, performance reflects the amount of material that can be simultaneously maintained in PM.

[2]Baddeley's (2000) model does allow for direct transfer of phonological information from PSTM to PLTM. It is this link that presumably underlies vocabulary learning. The issue here is whether the PSTM–PLTM link is also involved in learning sequence information or whether in that case the episodic buffer is required.

[3]Items that involved position errors were not scored as totally incorrect because this measure was intended to be more sensitive to the storage function of PSTM, rather than to the ability to retain serial position in random lists. Serial recall was required in order to prevent strategic recall of the most recent items.

[4]At the third generalization test the performance on the individual paradigms was as follows: Paradigm $1 = 69\%$, Paradigm $2 = 80\%$, Paradigm $3 = 66\%$, Paradigm $4 = 49\%$. A repeated-measures 2×2 ANOVA with gender (masculine, feminine) and regularity (Paradigms 1 and 2 versus Paradigms 3 and 4) as factors showed that there was an interaction between regularity and gender, $F(1, 19) = 8.58$, $p < .01$, $\eta^2 = .311$. The effect of regularity was confined to the feminine items. This pattern can be explained in terms of the operation of the general euphony bias mentioned in the description of the language (see "Method"). What is perhaps surprising is that performance on Paradigms 1 and 3 is so similar. This may reflect a negative euphony bias: a reluctance to use a determiner that ends in a different vowel from the noun. This would make *il libro* and *un letto* preferrable to *la libro* and *una letto*, and *il paese* and *un tegame* preferable to *la paese* and *una tegame*. The same negative bias would also explain the

poor performance for all of the items in Paradigm 4. Note, however, that although these general biases may have influenced generalization test performance, and presumably also learning of the system, we would argue that the patterns of individual differences are more a reflection of differences in learning item-specific constraints on determiner selection.

[5]Three participants fell into this category. Two participants achieved two consecutive 100% scores prior to Cycle 5.

[6]This low level of reliability presumably reflects the relatively small number of observations (24 as opposed to 48 for Input Memory in Experiment 1). Reliability might have been increased by including the trials from Cycle 2. However, this was not done, because it seemed likely that after the first generalization test there would be a greater probability that recall would be influenced by rule learning or conscious attempts at rule learning.

[7]The best 10 participants scored 76% correct on trained items, where chance was 50%, $t(9) = 14.45$, $p < .001$, whereas they scored only 54% on generalization items, $t(9) = 0.5$, *ns*.

References

Adams, A.-M., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. *International Journal of Psychology*, *34*, 364–373.

Adams, A.-M., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology*, *49A*(1), 216–233.

Adams, A.-M., & Willis, C. (2001). Language processing and working memory: A developmental perspective. In J. Andrade (Ed.), *Working memory in perspective* (pp. 79–100). Hove, England: Taylor & Francis.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Baddeley, A. (1990). *Human memory: Theory and practice*. Hove, England: Lawrence Erlbaum.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586–595.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*, 507–584.

Braine, M. D. S., Brody, R. E., Brooks, P. D., Sudhalter, V., Ross, J. E., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Langage*, *29*, 591–610.

Brooks, P. J., Braine, M. D. S., Catalano, L., & Brody, R. (1993). Acquisition of gender-like noun classes in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32*, 76–95.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, *39*, 535–594.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.

Caselli, M. C., Leonard, L. B., Volterra, V., & Campagnoli, M. G. (1993). Toward mastery of Italian morphology: A cross-sectional study. *Journal of Child Language*, *20*, 377–393.

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory and Cognition*, *21*, 162–167.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 671–684.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 450–466.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, *48*, 631–664.

Ellis, N. C., Lee, M. W., & Reber, A. R. (1999). *Phonological working memory in artificial language acquisition*. Manuscript in preparation, University of Wales, Bangor.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology*, *49A*(1), 234–250.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, England: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, *39*, 218–245.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73–100). Hove, England: Psychology Press.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2,* 103–127.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum.

Goldberg, A. E. (1998). Patterns of experience in patterns of language. In M. Tomasello (Ed.), *The new psychology of language* (pp. 203–219). Mahwah, NJ: Lawrence Erlbaum.

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition, 29,* 503–511.

Holmes, V. M., & de la Batie, B. D. (1999). Assignment of grammatical gender by native speakers and foreign language learners. *Applied Psycholinguistics*, *20*, 479–506.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, *21*, 421–452.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behaviour*, *10*, 409–416.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, England: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of context variable. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 158–169.

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

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.

Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, *21*, 557–587.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Working memory, interactional feedback and L2 development. In P. Robinson & P. Skehan (Eds.), *Individual differences and second language instruction* (pp. 151–209). Philadelphia: Benjamins.

Maratsos, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 240–266). Cambridge, England: Cambridge University Press.

Maratsos, M., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127–214). New York: Gardner Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J.-R., & Druhan, B. (1989). The role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1083–1100.

Nayak, N., Hansen, N., Krueger, N., & McLaughlin, B. (1990). Language-learning strategies in monolingual and multilingual adults. *Language Learning*, *40*, 221–244.

Page, M., & Henson, R. (2001). Computational models of short-term memory: Modelling serial recall of verbal material. In J. Andrade (Ed.), *Working memory in perspective* (pp. 177–198). Hove, England: Taylor & Francis.

Perruchet, P. (1994). Learning from complex rule-governed environments: On the proper functions of non-conscious and conscious processes. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and unconscious information processing* (pp. 811–835). Cambridge, MA: MIT Press.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129–191.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*, 45–99.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on incidental second language learning: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson & P. Skehan (Eds.), *Individual differences and second language instruction* (pp. 211–266). Philadelphia: Benjamins.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction*. New York: Cambridge University Press.

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*, 608–623.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *50A*, 216–252.

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental approaches to second language acquisition* (pp. 81–94). Oxford, England: Pergamon.

Skehan, P. (1998). *A cognitive approch to language learning*. Oxford, England: Oxford University Press.

Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language*, *13*, 69–91.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371–391.

Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics*, *20*, 119–148.

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology*, *52A*(2), 303–324.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*, 156–163.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.

Williams, J. N. (in press). Induction of noun classes by adult learners and connectionist models. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and the lexicon in second language acquisition*. Amsterdam: John Benjamins.