*SSLA,* **21**, 1–48. Printed in the United States of America.

# Memory, Attention, and Inductive Learning

John N. Williams
*University of Cambridge*

Three experiments investigated the relationship between memory for input and inductive learning of morphological rules relating to functional categories in a semiartificial form of Italian. A verbatim memory task was used as both the vehicle for presenting sentences and as a continuous measure of memory performance. Experiments 2 and 3 introduced increasingly explicit manipulations of attention to form compared to Experiment 1. In all experiments there were strong relationships between individual differences in memory for input as measured early in the experiment and eventual learning outcomes, and in Experiments 2 and 3 learning form-form (but not form-function) rules was related to vocabulary learning efficiency (taken as a measure of phonological long-term memory ability). These relationships along with the lack of an effect of feedback in Experiment 3 suggest that subjects tended to adopt a data-driven, as opposed to conceptually driven, mode of learning. However, the fact that the introduction of highlighting and vocabulary pretraining in Experiment 2 had a large impact on learning without improving early memory is taken to suggest that knowledge of distributional rules does not simply emerge out of memory encodings of the relevant forms but depends upon the appropriate allocation of attention over relationships between input elements at the time of encoding.

The debate over the relative merits of explicit instruction, that is, deduction, as opposed to self-directed rule discovery, or induction, has a long history in language teaching (see Decoo, 1996, and Hammerly, 1975, for reviews). Decoo

Many thanks to the anonymous *SSLA* reviewers for their helpful and constructive criticism of earlier drafts of this paper. Special thanks to Peter Robinson for very helpful advice concerning presentation and to Nick Ellis for his comments on an earlier report of some of the present experiments.

Address correspondence to John N. Williams, University of Cambridge, Research Centre for English and Applied Linguistics, Keynes House, Trumpington Street, Cambridge, CB2 1QA, UK; e-mail: jnw12@hermes.cam.ac.uk.

© 1999 Cambridge University Press 0272-2631/99 $9.50



https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

defines induction as "the process that goes from the specific to the general, namely first the real language use, from which will 'emerge' patterns and generalizations" (p. 96). The advantage of induction, according to Hammerly, is that the knowledge that is acquired is couched in terms that make sense to the individual student rather than in terms imposed by the instructor (in contrast to explicit instruction). In one form of the technique, students are encouraged to discover grammatical rules for themselves through inspection of target language sentences, often presented in contrastive pairs (Fischer, 1979). Thus, the emphasis is on self-directed and conscious rule discovery. Decoo (1996) refers to this type of induction as "conscious induction as guided discovery" (p. 96). The focus of the present study, however, is upon inductive learning from input that is not overtly structured, from whole sentences that are not juxtaposed so as to highlight specific grammatical points. Teaching techniques that eschew formal instruction, or overt manipulation of the input, must be assuming that generalizations will emerge from exposure to sentences in the target language, and it is the factors that influence this kind of induction that are under scrutiny here. Of particular concern will be the importance of memory and attention.

## A THEORETICAL FRAMEWORK FOR ADDRESSING THE RELATIONSHIP BETWEEN MEMORY, ATTENTION, AND INDUCTIVE LEARNING

Robinson (1995) provides a general overview of the psychology of memory and attention from an applied linguistic perspective. The first point to note is that there is a close, and necessary, relationship between conscious attention and encoding in long-term memory (see, for example, Cowan, 1993; Logan & Etherton, 1994; Schmidt, 1990). Only the material that is consciously attended enters long-term memory, a process that Schmidt refers to as "noticing." Not everything that is registered by the senses is therefore encoded in long-term memory, although it may activate long-term memory representations at quite a deep level (e.g., subliminally presented words may activate their meanings without the subject having any conscious awareness of them; Marcel, 1983). Thus, there is likely to be a very important connection between the way a particular task causes conscious attention to be distributed over the input and what is actually stored in memory (see Logan & Etherton, 1994, for an empirical demonstration of this point).

However, attention is not a sufficient condition for long-term memory encoding. Rather, long-term memory encoding depends on additional processes of what Robinson (1995) refers to as "rehearsal." In the case of input that is either presented in, or can be converted into, phonological form, representations in short-term memory can be cycled through the "phonological loop" (Baddeley, 1986) so that their activation level is periodically refreshed (Cowan, 1993). Following Craik and Lockhart (1972) and Robinson (1995), we may refer to this as "maintenance rehearsal," and characterize it as an essen-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

tially "data-driven" process. A second form of rehearsal involves establishing connections between the information in short-term memory and structures in long-term memory, a process that was implicit in the notion of "depth of processing" (Craik & Lockhart, 1972) but can be more generally regarded as an effect of elaboration, of embedding new information in old memory structures or schemata (see Baddeley, 1990, for discussion). Again, following Craik and Lockhart, and Robinson, we may refer to this process as "elaborative rehearsal" and characterize it as a "conceptually driven" process.

Both types of rehearsal can lead to learning, but in somewhat different ways. Maintenance rehearsal will lead to a long-term memory encoding of the input in an unanalyzed form. As representations of input accumulate in memory, inductive learning mechanisms might automatically extract generalizations. Although there is no agreement as to whether such mechanisms are best modeled using connectionist (see Davis, 1992, for an introduction), rule-based, or production system formalisms (Anderson, 1983; Holland, Holyoak, Nisbett, & Thagard, 1986), there is general agreement that at least some (if not all) aspects of language learning can be regarded as a process of breaking down initially unanalyzed chunks of language in long-term memory into smaller, more productive units (Bialystok & Ryan, 1985; Ellis, 1996; MacWhinney, 1982; Peters, 1983). This type of learning will be referred to here as *data-driven*.

According to Robinson (1995), elaborative rehearsal can be identified with nonautomatic learning processes that are engaged by a conscious intention to learn generalizations. The learner may consciously retrieve instances of previous language input and compare them with the present input in an attempt to derive a rule (see Ellis, 1994, for a discussion of this process in relation to vocabulary acquisition), or a previously formed hypothesis might be consciously retrieved and tested against the current input. In both cases, the learner can be regarded as elaborating the relationship between the current input and long-term memory representations. This type of learning will be referred to as *conceptually driven*. It is generally assumed to be more effortful than data-driven learning.

From the above, it is clear that data-driven and conceptually driven learning make somewhat different demands on memory. Whereas data-driven learning requires accurate long-term memory encodings of large amounts of prior input in unanalyzed form, conceptually driven learning relies on a combination of a much more selective sampling of the input data and previously entertained hypotheses. In the present experiments, the hallmark of data-driven learning will be assumed to be a correlation between inductive learning and the ability to encode the input data in long-term memory, especially as measured early in the experiment when it must presumably be stored in a relatively unanalyzed form. Only in this case can it be argued that memory ability plays a direct causal role in learning by determining the quality of long-term memory representations of input upon which data-driven inductive learning processes operate. Relationships between memory performance as measured

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

later in the experiment and subsequent measures of learning could reflect the influence of learning on memory performance (discussed further below) but do not indicate whether that learning occurred in a data-driven or conceptually driven mode.

## Implicit versus Explicit Learning

Implicit learning is said to occur when learners have no intention to learn (i.e., the learning process is automatic) and when they are unaware of what they have learned. Explicit learning refers to situations in which learners intend to learn and when they are aware of what they have learned. The distinction between data-driven and conceptually driven learning might seem at first sight to naturally align with that between implicit and explicit as characterizations of the learning process and the resulting knowledge. Indeed, Robinson (1995) argues, "Implicit test performance reflects study-test overlaps in data-driven processing, and explicit test performance reflects study-test overlaps in conceptually driven processing" (p. 311).

A great deal of research—both within experimental psychology (see Winter & Reber, 1994, for a review) and applied linguistics (De Graaff, 1997; DeKeyser, 1995; Ellis, 1993; Robinson, 1996, 1997a, 1997b)—has been directed at the issue of whether implicit learning is really possible. There is evidence that, under task conditions that are assumed not to encourage an explicit, conceptually driven mode of learning, subjects do seem to be able to form long-term memory representations of input material (or at least fragments thereof) and those representations do seem to inform subsequent grammaticality judgments on previously unseen items without the subjects' awareness. However, there is much debate as to whether what is learned implicitly can be properly characterized in terms of abstract, generalizable rules (Knowlton & Squire, 1994; Perruchet, 1994). Indeed, the evidence for implicit rule learning is almost nonexistent among studies that have used real, or natural-like, language systems (DeKeyser, 1995; Ellis, 1993; Robinson, 1996, 1997a, 1997b). From a reading of this literature, it might be concluded, therefore, that data-driven learning has no role in (adult) second language acquisition.

However, before accepting the above conclusion it should be borne in mind that there is not necessarily a direct correspondence between implicit and data-driven learning. Data-driven learning implies learning without intention (generalizations automatically emerge from long- term memory encodings of input), whereas implicit learning refers to learning not only without intention but also without awareness of the product of learning. As Schmidt (1994) points out, these two aspects of learning are, in principle, dissociable. Schmidt speculates that it may be possible to learn intentionally through a conceptually driven process of hypothesis formation and testing, but when the resulting rules have become automated, it may be possible to use them without awareness (in which case the process of learning is explicit, but the resulting knowledge is used implicitly). Of more relevance to the present ex-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

periments is an alternative form of the dissociation between intention and awareness that has been suggested by Ellis (1996): Learning may be supported by an automatic, nonintentional, data-driven process, but the learner might become aware of the resulting generalizations through the structuring effect that they have on perception of input. In this case, the process of learning is implicit, but the resulting knowledge becomes explicit.

The present experiments do not pretend to address the difficult issue of whether it is possible to learn natural language rules without awareness of either the process or product of learning. They address the more general question of the extent to which inductive learning can be characterized as data-driven as opposed to conceptually driven under various task conditions. Nevertheless, the task that the subjects were asked to perform did resemble the kind of memory task that has been traditionally used as an implicit learning task (Mathews et al., 1989; Reber, 1976; Robinson, 1996, 1997a, 1997b) in that the subjects were simply instructed to recall each input sentence as accurately as possible. A memory task was adopted because, first, it demanded maintenance rehearsal, thereby maximizing the chances of representations of input being encoded in long-term memory and leading to data-driven learning and, second, because it permitted the learners' memory for input to be continuously monitored, thereby allowing an exploration of the relationship between memory and learning.[1] As will become apparent, the measures of learning (the memory task itself and a subsequent translation task) were constructed so as to be potentially sensitive to either implicit or explicit knowledge. Debriefings showed that the subjects had a good deal of conscious knowledge of the target grammar, the accuracy of which was reflected in their performance. Even so, as argued above, it still makes sense to ask whether the learning process was data-driven or conceptually driven irrespective of subjects' awareness of the resulting knowledge.

## Individual Differences and Language Learning

The present experiments are concerned with the correlates of learning success, both at the macroscopic level of the impact of different modes of input presentation (which involves comparison of levels of learning between different groups of subjects who received the same material under different conditions) but also at the microscopic level of correlations between individuals' memory performance for input and their level of learning (and how these correlations are affected by variations in the input presentation mode). It is therefore relevant to consider the general issue of individual differences and learning and, in particular, research that has compared the correlates of learning under either implicit or explicit conditions.

It has been argued that implicit and explicit learning can be distinguished in terms of their relative dependence on individual differences in higher cognitive functions, such as those measured by intelligence tests, or measures of language aptitude such as the Modern Language Aptitude Test (Carroll & Sapon,

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

1959). Reber, Walkenfeld, and Hernstadt (1991) compared subjects' artificial grammar learning under implicit conditions (induced by a memory task) with their performance on a series-completion problem-solving task, which was assumed to involve explicit learning. They found that only learning in the explicit task correlated with individual differences in IQ. In the context of second language acquisition, Krashen (1985) claimed that only (explicit) learning is influenced by language aptitude. However, De Graaff (1997) found that learning outcomes under both instructed and noninstructed conditions correlated with aptitude (although the extent to which the noninstructed condition actually promoted the kind of learning referred to by Krashen as "acquisition" is not clear). On the other hand, Robinson (1997b) found that learning in an incidental condition (focus on meaning, corresponding more closely to Krashen's notion of an acquisition-rich environment) did not correlate with aptitude.

Although Reber et al. (1991) give the impression that implicit learning should be relatively immune to individual differences in general, at least in the case of learning natural languages, the above discussion of data-driven learning suggests that memory ability—specifically, the ability to encode novel and relatively unanalyzed forms in long-term memory—should influence the learning process because the accuracy of these representations will directly determine what generalizations can be formed. However, when Robinson (1997b) dissected aptitude into memory ability (as measured by the paired-associates subtest on the MLAT) and grammatical sensitivity (as measured by the words in sentences subtest), he found that memory only correlated with learning in the rule-search and instructed conditions of the experiment (i.e., the conditions that promoted explicit, conceptually driven learning). There were no correlations in the implicit (memory task) and incidental (focus on meaning) conditions. On the other hand, Ellis and Schmidt (1997) performed an artificial-grammar learning experiment in which a memory task was used to induce an implicit learning mode and found correlations between both short- and long-term memory for input sentences and learning, but only in the case of relatively long-distance dependencies (agreement between nonadjacent elements). They interpret this result as showing that short-term memory capacity determines the length of the string that can be transferred to long-term memory, which is clearly important for learning long-distance dependencies. However, from the perspective of data-driven learning, one might still expect to find some relationship between recall accuracy for input strings and subsequent learning for short-distance dependencies, unless these were so simple as to be learned even by subjects with poor memory (which was not the case because overall levels of learning were extremely low in all conditions). Furthermore, in order to establish a direct connection between memory and learning, it is necessary to examine how learning of specific structures is related to memory for the relevant aspects of form from early on in the experiment. Ellis and Schmidt (1997) only report correlations involving general levels of recall accuracy, and the significant correlation between memory and learning for long-distance dependencies is not broken down according to the

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

point in the experiment at which memory performance was measured. In the present experiments, memory for specific aspects of form was measured at different points and related to ultimate levels of learning of the relevant rules.

## Subsidiary Issues

Apart from an examination of the relationship between memory and ultimate levels of learning, two subsidiary issues were addressed by the experiments. First, would memory for input sentences improve over the course of the experiment and could that improvement be attributed to learning grammatical generalizations, as opposed to increasing familiarity with the phonology, lexis, and grammatical morphemes of the language? This is a test of Ellis's (1996) view of language learning as a "bootstrapping" process whereby memory and learning continually reinforce each other. In work on artificial grammar learning, improvements in memory in the course of the learning phase have often been reported (e.g., Mathews et al., 1989; Reber et al., 1991). On the other hand, in the case of language learning, Bley-Vroman and Chaudron (1994) argue that it is possible that subjects will be unable to utilize grammatical knowledge with sufficient automaticity to aid recall, especially in the early stages of learning (hence casting doubt on elicited imitation as a means of investigating the development of grammatical knowledge).

Second, the present target language incorporated form-form mappings (i.e., concord: article-noun and modifier-noun agreement) as well as form-function mappings (i.e., grammatical functions: verb inflections marked person and number; noun inflections marked gender and number). Laboratory studies of grammar learning have largely neglected form-function mappings. One notable exception is a study by DeKeyser (1995). He obtained no evidence of implicit learning of form-function mappings, despite the fact that some of the rules were very simple (e.g., the addition of *-on* to mark plurality of a noun) and despite providing large amounts of exposure (2,480 trials). DeKeyser used a task that was essentially meaning-oriented (sentence-picture matching in which mismatches concerned errors of vocabulary, not grammar). It may be thought that a verbatim memory task, being oriented to surface form, might be even less likely to result in learning of form-function mappings, especially given arguments that learners face a basic problem in simultaneously attending to form and meaning (VanPatten, 1990). On the other hand, learners might find it easier to incidentally encode meaning in the context of a form-oriented task than to incidentally encode form in the context of a meaning-oriented task. If encoding of form in long-term memory requires more effort and is less automatic than encoding of meaning, then in a meaning-oriented task there may be little incidental encoding of form, but in a form-oriented task there might be a greater likelihood of incidental encoding of meaning. Alternatively, subjects in the present experiments might perceive the level of meaning as what Bley-Vroman and Chaudron (1994) refer to as a relevant "level of control" at which to perform the recall task; they might pay attention to form-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

function correspondences because they realize that meaning can serve as a cue for retrieving form. Therefore, it is of some interest to see whether a verbatim memory task can lead to learning of functions and, if so, whether learning form-function mappings shows the same relationship to memory as form-form mappings.

## Overview of the Tasks Used in the Present Experiments

In the present experiments, a verbatim memory task served both as the vehicle for presenting language input and as a continuous measure of memory performance. Subjects were presented with sentences in a language (Italian) with which they were not previously familiar. Each sentence had to be recalled as accurately as possible after a filled interval. This task was presented to the subjects as a straightforward exercise in memorizing sentences and no mention was made of any subsequent tests of grammar. After performance of the memory task upon 40 sentences, the subjects' knowledge of grammatical rules was evaluated in an English-to-Italian translation task that required the selection of Italian words from an array of possibilities. This will be referred to as the translation task (see Experiment 1, Procedures for the rationale behind the design of this task). It was then possible to relate levels of learning to recall performance as measured at various points in the memory task. Experiments 2 and 3 introduced modifications that, it was hoped, would enhance learning. In Experiment 2, the vocabulary was pretaught and the grammatical morphemes were highlighted. In Experiment 3, feedback was provided during recall attempts in the memory task.

As argued above, the memory task might also provide a measure of learning, as well as a measure of memory. However, distinguishing improvements in memory performance that are due to learning of the specific form-form and form-function mappings under study is not straightforward. The obvious strategy would be to compare memory performance on sets of sentences at the beginning and end of the memory task, taking care to use unique lexis in the critical sentences (this would be analogous to the use of transfer tests in previous research on miniature language learning, in which exponents of the same underlying rule system, different from those appearing in training, are used to test for generalizable knowledge as opposed to memory for the training stimuli). In the present case, however, memory for sentences could improve because of increased familiarity with the general sound structure of the language, word-order patterns (which were not of concern here because they appeared to be learned quite readily), and the forms of those morphemes that necessarily occur in both the training and test sentences (specifically, articles and stems of possessive adjectives). Therefore, in order to evaluate the extent to which any improvement in memory for test sentences at the beginning and end of the experiment was due to learning the target rules, Experiment 1 also contained a control condition in which the input consisted of a corrupted form of Italian. The corrupted Italian was constructed in such a way as to pre-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

vent generalizations from being formed, while maintaining exactly the same frequency of exposure of all grammatical and nongrammatical morphemes as in the real Italian condition. Comparisons between these two groups in terms of improvements in accuracy and increases in grammaticality of recall provided a measure of sensitivity to the target rules during the memory task. This condition also provided a baseline against which to evaluate performance in the translation task (given that choice behavior in this task might have deviated from chance levels for reasons unrelated to the material presented in the memory task, such as prior knowledge of other languages).

### Summary of Research Questions

***Research Question 1: Memory as a Determinant of Inductive Learning.*** Will individual differences in levels of performance in the translation task correlate with memory for input sentences as measured early in the experiment? If so, learning might be regarded as data-driven. If not, variations in memory for input can be assumed to be of minor importance to inductive learning compared with other sources of individual differences.

***Research Question 2: Memory as a Consequence of Learning.*** In the memory task, will there be greater memory gains in the real Italian condition than in the corrupted Italian condition? This would indicate an effect of learning grammatical generalizations on recall. One might also expect to find correlations between translation-task performance and final levels of memory performance in the real Italian condition.

***Research Question 3: Induction of Functions.*** Will evidence of learning provided by the translation task be confined to form-form mappings, or will form-function mappings also be learned, and, if so, will the relationship between learning and memory be the same in the two cases?

## EXPERIMENT 1

### Subjects

A total of 26 subjects participated in the experiment and were paid £10.00 for participating. They were all Cambridge University students, average age 22 years, studying sciences, 20 of them at undergraduate level and 6 at postgraduate level.[2] All but two of them had taken language exams for the General Certificate of Secondary Education at age 16. Fourteen had taken French, four German, and six had taken both French and German. Six subjects expressed having some knowledge of one other language (Spanish, Greek, Russian, or Irish, and in two cases Latin), but in most cases they classed their knowledge as very basic, or survival level. Twelve of the subjects had used a foreign language outside a formal educational context and this was an average of 3.4 years prior to the experiment. A total of 16 subjects were tested on real Italian

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

(10 received real A, and 6 received real B; see below for details of the difference between these two versions). A further 10 subjects received corrupted Italian.

## Method

***The Language.*** A highly regular subset of Italian grammar rules was used to generate the sentences used in the experiments. There were basically two types of sentence structure (the bracketed element is optional):

| | | | | | |
|---|---|---|---|---|---|
| (1) | (Time adverbial) | + Verb | + Definite article | + Noun | + Adjective |
| | *Ogni weekend* | *studiano*-3.pl. | *la*-f.sg. | *storia*-f.sg. | *italiana*-f.sg. |
| | every weekend | study | the | history | Italian |
| | "Every weekend they study Italian history." | | | | |

| | | | | | |
|---|---|---|---|---|---|
| (2) | (Time adverbial) | + Verb | + Definite article | + Possessive adjective | + Noun |
| | *Spesso* | *riparo*-1.sg. | *la*-f.sg. | *mia*-f.sg. | *macchina*-f.sg. |
| | often | repair | the | my | car |
| | "Often I repair my car." | | | | |

There were no overt subjects, the subject being marked by the inflection on the verb (in the full grammar subjects are merely optional). Only the present simple tense was used. First-person singular was marked by the inflection *-o*, third-person singular by *-a*, first-person plural by *-iamo*, third-person plural by *-ano* (this is only one of a number of inflectional paradigms for the present tense in the full grammar). The second person was not used. Definite articles were obligatory. Adjectives always came after the nouns (in the full grammar, some adjectives can come before) and articles came before possessive adjectives. Articles, adjectives, possessive adjectives, and nouns all agreed in gender and number. Only regular nouns and adjectives were used, for which the masculine singular ending was *-o*, masculine plural ending was *-i*, feminine singular ending was *-a*, and feminine plural ending was *-e*. Therefore, agreement in this case reduced to euphony—that is, the same ending (this is not the case in the full grammar, in which gender and number cannot be predicted from the noun ending). The four articles were *il* for masculine singular, *i* for masculine plural, *la* for feminine singular, and *le* for feminine plural (in the full grammar, other definite articles are possible depending on the onset of the noun). Note that, in three of these four cases, noun agreement again reduced to euphony between endings (e.g., *i libri*, *la donna*, *le donne*), the exception being *il* (e.g., *il libro*). Four types of possessive adjective were used: *mio*/*miei*/*mia*/*mie* ("my"), *suo*/*suoi*/*sua*/*sue* ("his/her"), *nostro*/*nostri*/*nostra*/*nostre* ("our"), and *loro* ("their"). Again for three of the types, agreement with the object noun reduced to euphony. The exception was *loro*, which never changed.

This language permitted an examination of learning of the following, potentially dissociable, types of knowledge: segmentation into stems and affixes,

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

form-form relations (noun-modifier and article-noun agreement), and form-function relations (functions of noun and verb inflections). Note that the syntax of the language deviated from that of English in other significant respects: There were no overt subjects, adjectives followed nouns, and articles appeared in possessive noun phrases. However, given that subjects generally seemed to have little difficulty with these aspects of syntax, the analysis will concentrate upon learning of morphological rules relating to functional categories.

***Design and Materials.*** In the experiment, subjects performed a verbatim memory task upon sentences generated by the above grammar. The memory task was divided into five blocks that comprised eight sentences each. The comparison between Blocks 1 and 5 constituted the measure of improvements in memory performance, and the sentences that they contained will be referred to as test sentences. Each of the 16 test sentences used for Blocks 1 and 5 contained unique vocabulary. For the real Italian condition, each grammar point occurred with equal frequency in each block (e.g., there were two exponents of a first-person singular verb in each block, although all of the verb stems were different). In order that comparisons between Blocks 1 and 5 not be confounded with item difficulty, two sets of test items were rotated around the two blocks between subjects. That is, half of the subjects received one set of eight test sentences in Block 1 and another set of eight test sentences in Block 5, and these assignments were reversed for the other half of the subjects.

The 24 sentences presented in Blocks 2–4 of the memory task will be referred to as training sentences. These sentences were composed using four different verbs, four adjectives, and 24 nouns. Because the verb and adjective stems occurred more than once, bearing different inflections, there was in principle enough information across the input sentences to induce the underlying rules. Each noun appeared only once, although each possible inflection occurred equally often (this may not have been conducive to learning the functions of the inflections, see below).

In the corrupted Italian condition, the sentences were paired with the same meanings as the real Italian sentences from Blocks 1 to 4, but the grammatical morphemes were distributed in such a way as to prevent any generalizations being formed. Across different sentences, the verb and adjective stems always occurred with the same inflection. Particular verbs occurred with subjects of all possible types, and nouns rarely agreed with their accompanying articles or modifiers (e.g., *Amiamo il vini italiane* for "They love Italian wine," *Ogni sera studia le nostro testi* for "Every evening I study my textbooks"). Thus, there was no basis for inducing the inflectional status of the adjective and verb endings nor their functions or agreement constraints. For the test sentences, the sentences that appeared in Block 1 occurred in a corrupted form, whereas those in Block 5 appeared as real Italian. These subjects therefore received four blocks of corrupted Italian and shifted to real Italian in the final block.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

Further details of the sentence composition appear in the Appendix along with a complete listing of the items used in the memory task in their order of presentation. In general, an attempt was made to select words that would not pose particular pronunciation problems for the subjects. Additionally, for the training sentences of the memory phase, an attempt was made to reduce the memory load by selecting as many cognate items as possible. In order to satisfy the requirements on the frequency of inflections and verb and adjective stems, some of the sentences sounded rather contrived. A short English context was therefore written for each sentence in order to increase its plausibility (e.g., context: My wife is a researcher for a newspaper. Input sentence: *Studia i politici famosi*, "She studies famous politicians"). Also, there were two practice sentences. These were generated by the same grammar as for the training sentences, although unique nouns and verbs were used.

After the memory phase was complete, there followed a translation task in which subjects translated novel English sentences by selecting items from a visual array. There were 24 sentences divided into three blocks of eight sentences each. Given the difficulty of finding suitable verbs, the verbs used in each block were based on the verbs used in each set of test items (but this time the correct answer required that they carry a different inflection). It was also necessary to use forms of the possessive adjectives that had already occurred in the memorization task. To minimize familiarity effects, the items used in Block 1 of the translation task used the same verb stems and possessive adjective forms that the subject had received in Block 1 of the memorization task. Because of the rotation of items between Blocks 1 and 5 of the memorization task, this meant that half of the subjects received one set of items in Block 1 of the translation test and another set of items in Block 2, and for the other half of the subjects this assignment was reversed. Block 3 of the translation task utilized completely unique lexical items, however.

After running the real and corrupted Italian conditions of the experiment, it was found that subjects had remarkable difficulty learning the article and noun systems. In the case of the nouns, it was felt that the fact that in the presentation phase each noun had only been presented once would not have facilitated induction of the relevant inflections for marking gender and number. In the case of the articles, the English versions of the sentences used in the experiment did not contain articles (this was because the use of the present simple tense was most felicitous in sentences expressing habitual actions, which in turn tend to contain generic, rather than definite, reference). Obviously, the lack of an English correspondent for the Italian definite article could well have impeded its acquisition. New versions of the real Italian training sentences were therefore written in which each noun now occurred six times (twice with each singular and plural inflection) and the English translations of the sentences now contained definite articles (the contexts for the meaning presentation were altered to make this possible). These materials will be referred to as the real B sentences, and those from the first version of the experiment the real A sentences. After testing six subjects in this phase, it was

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

found that there were no improvements in learning despite these changes. The data from the subjects who received real A and real B Italian were therefore pooled. The greater number of subjects in the pooled real Italian condition allowed comparison between subsets of the real Italian subjects and the corrupted Italian subjects, as well as further exploration of the basis for individual differences in learning.

***Procedures.*** Subjects were recruited through a bulk mailing directed to students at Cambridge University. The letters described the experiment as concerning people's memory and processing of unfamiliar languages, and this information was repeated to volunteers before the experiment began. The subjects then filled in a short questionnaire concerning their language background. They were given instructions and practice in the memory task, and no mention was made of there being any subsequent test of grammar learning. Subjects were tested individually in a small room set aside for this purpose. The whole testing session took between 1 hour 40 minutes and 2 hours. The experiment was entirely computer-controlled, and the material was developed using the authoring package Macromedia Authorware Professional version 3.5 running on a Power Macintosh.

*Memory task.* The meanings of the Italian sentences were presented using semigraphical representations, two examples of which appear in Figure 1. Subjects were given practice at interpreting such figures before the experiment began (none of the practice items occurred in the experiment). The subjects were told to read the context and to interpret the figure in English, saying the English sentence aloud. This was to encourage deeper processing of the meaning of the sentence than might have been the case if the English sentence had simply been provided in written form. It was explained that the purpose of the context was to aid in the interpretation of the figure. Semigraphical representations were used in order to permit transparent presentation of the relationship between the Italian sentence and its meaning (see Figure 1). So that the spatial layout of the figures would not provide cues to word order, the arrangement of the components was varied.

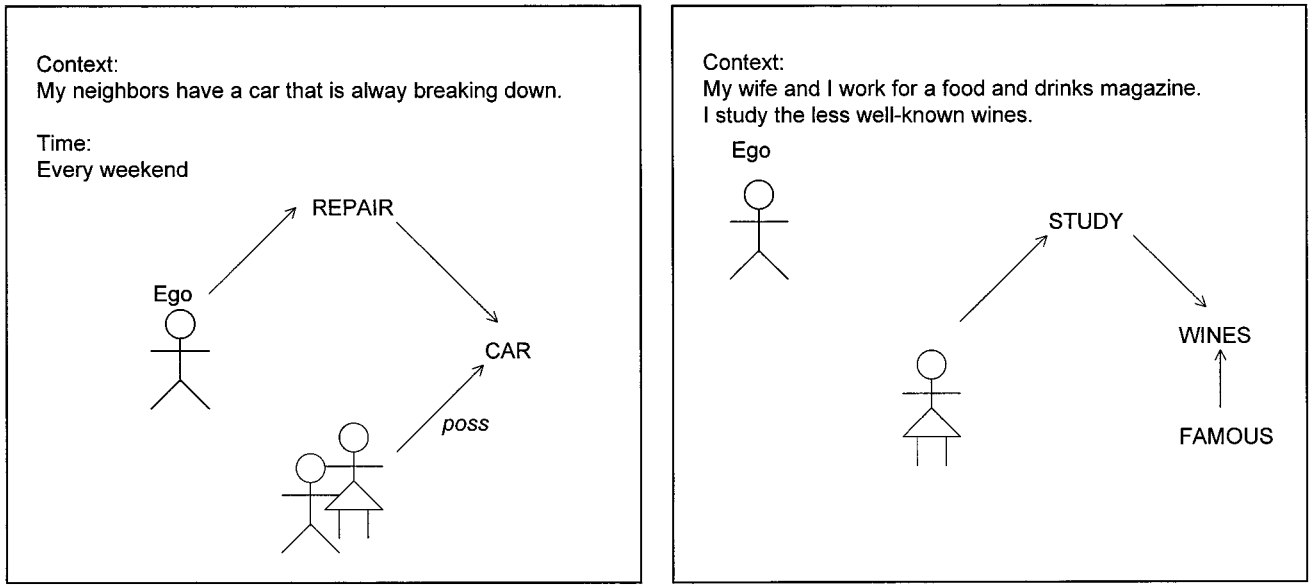
For the sentence presentation phase of the memory task itself the sequence of events was as follows:

1. A semigraphical meaning representation was presented and the subjects said the English sentence aloud.
2. The subjects clicked a "continue" button and the correct English sentence was displayed (this was to ensure that all subjects did in fact derive the correct meaning).
3. On clicking "continue" again, the equivalent Italian sentence was presented simultaneously both visually and auditorily (recorded by a female native Italian speaker). The written form remained on the screen for 2.5 seconds after the end of the spoken form while the subjects repeated and read it aloud.
4. The written form of the sentence disappeared. The subjects heard the sentence again one segment at a time and, as each segment was heard, a part of the semantic figure was highlighted in red. For the sentences containing adjectives the seg-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press






"Every weekend I repair their car"

"She studies the famous wines"

Figure 1. Examples of meaning representations ("poss" in the left-hand figure signifies possession).

ments consisted of the time adverbial (the English equivalent was highlighted), the verb (the stick figure, English verb, and arrow connecting them were highlighted), the article plus noun (the English equivalent and the arrow connecting it to the verb were highlighted), and the adjective (the English equivalent and the arrow connecting it to the noun were highlighted). The sequence for the sentences containing possessive adjectives was similar except this time the article and possessive adjective comprised a segment (the stick figures doing the possession and the arrow connecting them to the object noun were highlighted). After each segment was heard the subjects repeated it aloud, although they were only provided 2 seconds in which to do so.

5. The whole sentence was again presented simultaneously in written and spoken form, and the written form remained on the screen for 3 seconds while the subjects repeated it. After a delay of 3 seconds, the next trial began.

Given the timing of the above events, it was assumed that there was very little opportunity during the presentation sequence for the subjects to analyze the input sentences. They were either engaged in listening or repeating.

After every two sentences had been presented, the subjects entered a recall phase in which the first sentence was orally recalled, followed by the second. The sequence of events on a recall trial was as follows:

1. A semigraphical meaning representation corresponding to the meaning of the to-be-recalled sentence was displayed with the word "Recall" presented above it. The spatial layout of the figure was different from that which accompanied the original sentence. The subjects attempted to orally recall as much of the original sentence as possible.
2. On pressing "continue," the first letter of each content word was presented in red near the corresponding English word in the figure (no clues were provided for articles or possessive adjectives). On pressing "continue" again, more of the content word was revealed and, on pressing a third time, still more. At this third level of help the stem of each content word was presented. Thus, although subjects were given a large amount of help in recall, they were never provided with the grammatical morphemes.[3] They were instructed to retrieve as much of the Italian sentence as they could with no help and, as more help was provided, to alter their response if they wished to (however, they were encouraged not to reiterate the same response at each level of help).
3. On pressing "continue," the meaning representation for the second sentence to be recalled was presented. Note that subjects were not provided with the correct answer. This was so that they would not be encouraged to analyze the sentences.

After performing the above sequence on two practice sentences, the subjects cycled through the presentation and recall phases for the 40 sentences of Blocks 1–5 of the memory task. There was a minimum 40-second rest between blocks. This phase of the procedure was audiotaped.

*Translation task.* The sequence of events on each trial of the translation task was as follows:

1. A meaning representation was presented as in the memory phase. As before, subjects first produced the English aloud and, by pressing "continue," checked their

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

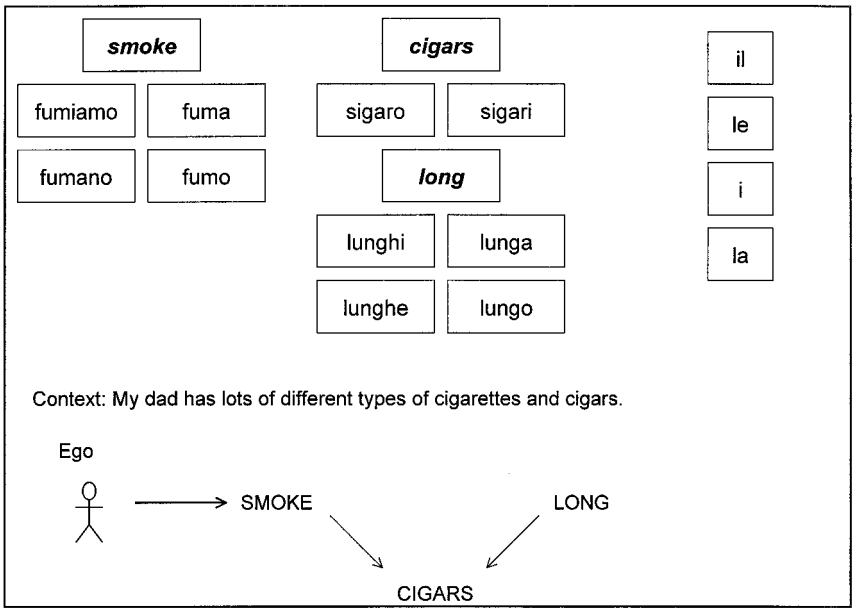



**Figure 2.** An example of a translation trial.

response against the correct one. Because any learning of the grammar had occurred in the context of these meaning representations, it was thought preferable to maintain the same format in the translation task.

2. On clicking "continue," the context and English translation disappeared and the display for the translation task for that trial was presented. An example appears in Figure 2. Arrangement of the possibilities for verbs, nouns, and adjectives was randomized so that the subjects could not use spatial information to help them understand the paradigm.

The subjects simply composed a response by clicking on the words required, which then appeared in a large blue rectangle. The computer recorded which selections were made. If at any point they wanted to change their response, they could start all over again by clicking on "try again." Otherwise they clicked on "done" and the next item was presented. Subjects were requested to think aloud as they performed this task, explaining the rationale for their choices as much as possible or indicating whether they were guessing.

The translation task was designed to provide a sensitive measure of the subjects' knowledge of both form-form and form-function relations in the context of new vocabulary. It tapped the ability to make discriminations between the alternatives provided, either on the basis of form-level or meaning-level factors. In one sense, this task can be regarded as tapping the kinds of selec-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

tion processes that operate in normal production, except that here the alternatives are provided on the screen rather than retrieved from memory (this was intentional because the aim was to assess grammatical knowledge without the additional problems that might be imposed by retrieval of the forms themselves). However, it cannot be assumed that the alternatives provided on the screen correspond to the alternatives available in memory for a native speaker of Italian (for example, representations of verbs might be decomposed into stems and affixes for the native speaker, whereas in the present task they were presented as whole words).[4] It should also be noted that, even though there is a sense in which the translation task assesses productive knowledge, there is no reason to suppose that such knowledge is necessarily explicit. Subjects' choices could, in principle, be informed by unconscious knowledge, leading to relatively good performance even though they may have the subjective impression of guessing.

Block 1 of the translation task provides the critical measure of learning for the analyses reported below. For Blocks 2 and 3, feedback was given. When the subjects clicked on "done," they received the correct answer in another rectangle directly below their own response. They also heard the sentence. Two seconds after the offset of the spoken sentence, the correct written answer disappeared and they were instructed to change their own answer accordingly from memory using the "try again" button. They could keep cycling though this sequence until they had produced the correct answer. Subjects were not required to think aloud during Blocks 2 and 3. The purpose of introducing feedback was to gauge whether the various aspects of the grammar were learnable when an explicit hypothesis-testing mode of learning was encouraged. It also permitted the subjects to evaluate how much they had learned and, it was hoped, to achieve a reasonable level of accuracy before finishing the experiment.

***Scoring Methods.***

*Memory task.* Subjects' recall attempts in Blocks 1 and 5 were scored for accuracy (with respect to the target sentence) of the following form types: verb inflection ($n = 8$), noun inflection ($n = 8$), article ($n = 8$), adjective inflection ($n = 4$), possessive adjective inflection ($n = 3$ because *loro* does not inflect). In order to score as correct, it was only necessary for the form to be produced as part of any of the successive recall attempts for that item. This very lax scoring procedure was used because it was not possible to decide in any principled way which of the recall attempts should be given precedence (in early recall attempts, inflections might not be produced simply because the relevant stem cannot be recalled, and in later attempts, recall might be corrupted through overcorrection of earlier recall attempts, or through interference from subsequently recalled forms; therefore, an accurately recalled grammatical morpheme in any recall attempt should be given credit). The articles, adjectives, and possessive adjectives were also scored for agreement with the accompanying noun in the recall attempt, regardless of correctness

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 1.** Mean percent correct recall in Experiment 1

| Condition | Block | Verb Inflection | Noun Inflection | Article | Adjective Inflection | Possessive Inflection |
|---|---|---|---|---|---|---|
| Real Italian | 1 | 47 | 76 | 33 (27) | 77 (64) | 54 (48) |
| A + B | 5 | 86 | 84 | 55 (47) | 83 (83) | 60 (58) |
| | (Gain) | +39* | +8 | +22*(+20*) | +6 (+19*) | +6 (+10) |
| Corrupted | 1 | 47 | 60 | 26 | 70 | 23 |
| Italian | 5 | 61 | 71 | 44 (40) | 65 (65) | 33 (27) |
| | (Gain) | +14* | +11* | +18* | –5 | +10 |

*Note.* *$p < 0.05$. Percentage agreement with the recalled noun is shown in parentheses. Note that the agreement score is irrelevant for corrupted Italian.

with respect to the original sentence (although cases where the article or modifier was forced to agree with a misrecalled noun were rare in this experiment). The relevant items had to be part of the same recall attempt but, again, any recall attempt was valid. This method of scoring was not possible in Block 1 of the corrupted Italian condition.

*Translation task.* Subjects' performance was scored in terms of accuracy for the verbs ($n = 8$) and nouns ($n = 8$), which was taken as an indication of knowledge of the functions of the relevant inflections. Scoring was in terms of agreement with the noun for articles ($n = 8$), adjectives ($n = 4$), and possessive adjectives ($n = 3$ because *loro* does not agree). An extra category "possessive type" ($n = 4$) was also included to indicate whether subjects chose a possessive adjective of the correct type regardless of the inflection. As for the memory task, each of these categories will be referred to as form types. Word order errors were not taken into consideration in the present analyses. Occasionally an article was inserted as an overt subject and, in such cases, it was not counted.

## Results

***Memory Task.*** The changes in the materials for Blocks 2–4 between real A and real B made very little difference to the pattern of gains between Blocks 1 and 5 for articles and nouns. Therefore, the data for real A and real B was pooled for comparison with the corrupted condition. The mean percentage correct recall for each form type in Blocks 1 and 5 for real and corrupted Italian subjects is shown in Table 1, along with the percentage of articles, adjective, and modifier inflections that agreed with the accompanying noun. Two-tailed *t* tests were performed on the difference between Block 1 and Block 5 scores for each structure. Significant differences in these gains are indicated in Table 1.

The data for percent correct recall were submitted to an analysis of variance in which condition (real or corrupted Italian) was a between-subjects factor, and form type and blocks were within-subjects factors. The interaction between condition and blocks was not significant, $F(1, 24) = 1.22$. Averaged

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

over form types, memory gains between Blocks 1 and 5 were equivalent in the real and corrupted Italian conditions. Analyses of variance performed on individual form types showed that only in the case of verbs was there a significant interaction between condition and blocks, $F(1, 24) = 9.98$, $p < 0.01$, which was due to the fact that the memory gain was greater in the real Italian condition than in the corrupted Italian condition.

Scoring the articles, adjectives, and possessives for agreement with the accompanying noun in the recall attempt made little difference except in the case of adjectives. On this measure, the 19% gain in agreement between Blocks 1 and 5 was significant. It can also be seen from Table 1 that, for adjectives and possessives, Block 5 agreement scores were numerically higher in the real Italian condition than in the corrupted Italian condition. Two-tailed *t* tests on unrelated samples showed that this difference was significant for possessives ($p < 0.05$), but not for adjectives. When the adjective and possessive data were pooled to give a more reliable composite measure of noun-modifier agreement, then the difference between the two conditions was also significant ($p < 0.01$).[5]

The overall analysis of variance on the data from the real and corrupted conditions revealed a number of other significant effects. When averaged over condition and form type, recall improved between Blocks 1 and 5, $F(1, 24) = 44.9$, $p < 0.001$. Presumably this reflects an increased familiarity with individual grammatical morphemes (note in particular the significant improvement in recall for articles in both the real and corrupted conditions shown in Table 1) or more general factors such as familiarity with the sound structure of the language. Averaging over condition and blocks, the level of recall was found to differ according to form type, $F(4, 96) = 25.52$, $p < 0.001$. Recall of articles and possessive adjectives was particularly poor, but this is to be expected given that subjects were not given any recall cues for these forms. Also, averaging over condition, memory gains were not equal for all form types, as shown by an interaction between form type and blocks, $F(4, 96) = 6.37$, $p < 0.001$. In general, forms for which recall was already very good in Block 1 showed the smallest gain in Block 5, but this presumably merely reflects a ceiling effect.

***Translation Task.*** Once again, the changes in the materials between the real A and real B conditions made very little difference to subjects' performance on the articles and nouns, and so the data have been pooled. The percentages in each condition for the subjects in the real memory task are shown in Table 2, along with the difference in score with respect to the subjects in the corrupted memory task condition. Note that performance in Block 1, where no feedback was given, provides the best indication of learning in the memory task.

An analysis of variance was performed upon the translation data in which condition (real vs. corrupted memory condition) was a between-subjects factor, whereas form type (verbs, articles, etc.) and blocks were within-subjects

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 2.** Translation performance (%) in each condition of Experiment 1

| | Verb Function | | Noun Function | | Article-Noun Agr. | | Adjective-Noun Agr. | | Possessive-Noun Agr. | | Possessive Type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Block | Real | Diff | Real | Diff | Real | Diff | Real | Diff | Real | Diff | Real | Diff |
| 1 | 48 | +17 | 63 | +8 | 33 | +2 | 72 | +32* | 60 | +27 | 91 | +31* |
| 2 | 52 | +16 | 61 | +6 | 41 | +12 | 84 | +29* | 62 | +15 | 72 | +25* |
| 3 | 79 | +8 | 66 | +16 | 47 | +13 | 89 | +24* | 73 | +30* | 87 | +32* |
| (Mean) | 60 | +14 | 63 | +10 | 40 | +9 | 82 | +29* | 65 | +24* | 83 | +29* |

*Note.* Real = real A + B pooled data, Diff = real minus corrupted, *$p < 0.05$, Agr. = agreement.

factors. There was a main effect of subject group, $F(1, 24) = 11.40$, $p < 0.01$, because real-condition subjects performed better than the corrupted-condition subjects. However, it is clear from Table 2 that the difference between real and corrupted memory condition subjects was not uniform over form types. Table 2 shows the significance level of the difference between performance of the subjects in the real and corrupted memory conditions for each form type in each block, as calculated by individual analyses of variance. It can be seen that only in the case of possessive type, adjective-noun agreement, and to some extent possessive-noun agreement were the real memory condition subjects better than the corrupted-condition subjects. There was a notable lack of learning of the functions of the verb and noun inflections.

In general, the provision of feedback in Blocks 2 and 3 caused an improvement in performance. The overall analysis of variance showed that the main effect of blocks was significant, $F(2, 48) = 14.55$, $p < 0.001$. However, there was a significant interaction between form type and blocks, $F(10, 240) = 4.75$, $p < 0.001$, which indicates that the improvement was not uniform over form types. In particular, the categories of possessive type and noun showed no effect of feedback. There was no interaction between condition, form type, and blocks, $F < 1.0$.

The lack of significant learning of the functions of the verb inflections in the real condition contrasts with the significant improvements in memory for the inflections themselves in the memory task. This pattern would seem to suggest that the subjects had gained some knowledge of the verb system, which was apparent in their memory performance but not in their translation performance. As a further test of this, the 10 real-condition subjects with the worst performance on the translation task for verbs were selected and their memory task performance on verb inflections compared to the 10 subjects in the corrupted condition. This resulted in two groups who, as it happened, were matched exactly for mean translation-task performance on verbs (31% in each case).[6] A difference between the two subject groups in gain score for verbs in the memory task was still obtained—for the real-condition matched subgroup, memory for verb inflections rose from 34% in Block 1 to 80% in Block 5. An analysis of variance on the verb data alone showed that the interaction between subject group (real vs. corrupted) and block was significant,

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 3.** Correlations between translation and memory performance in Experiment 1

| Translation Task Performance | Block 1 Memory | Block 5 Memory |
|---|---|---|
| Modifier-noun agreement | .631** (.469) | .536* (.623**) |
| Article-noun agreement | .191 (.224) | .167 (.220) |
| Noun function | –.307 | –.111 |
| Verb function | .896** | .566* |
| (Overall) | .675** | .657** |

*Note.* *$p < 0.05$, **$p < 0.01$. Correlations when memory performance scored for agreement are shown in parentheses.

$F(1, 18) = 16.44$, $p < 0.001$. Therefore, this subset of subjects showed significantly greater improvements in memory for verb inflections than the corrupted-condition subjects but exhibited no greater knowledge of the functions of the inflections when tested in the translation task.

***Relationships between Memory and Translation Task Performance.*** In the real Italian condition, Block 1 translation-task performance on specific aspects of the grammar (as measured by the advantage over the corrupted Italian condition) was related to the general level of memory performance for the relevant structures in the memory task, especially as measured at the beginning of the experiment (i.e., in Block 1 of the memory task). Ignoring nouns for the moment, in Table 1 the rank order of Block 1 memory scores in the real Italian condition is adjective > possessive > verb > article, and in Table 2 the rank order of difference scores is adjective-noun agreement > possessive type > possessive-noun agreement > verb correct > article correct. It appears that subjects learned about the structures that they recalled best even as measured at the start of the experiment.

The same relationship between Block 1 memory and Block 1 translation was evident when individual differences were considered. In fact, there was a remarkable degree of variability between subjects' overall performance in Block 1 of the memory task and rather less variability by Block 5 (which presumably reflects a convergence in scores due to learning). For Block 1 memory, the mean was 55%, $SD = 14.8$, range 32–90. For Block 5 memory, $M = 75\%$, $SD = 10.6$, range = 55–97. Moreover, contrary to what one might expect, the variation in Block 1 translation was similar to that found in Block 1 memory, rather than Block 5 memory: $M = 55\%$, $SD = 17.4$, range = 32–87. Correlations between Block 1 translation-task performance and either Block 1 or Block 5 memory for the relevant inflections were computed on a form-type by form-type basis (although adjectives and possessives were combined into a modifier category to increase the reliability of the analysis, and *loro* was excluded because it is not inflected). Memory performance for articles and modifiers was scored both in terms of accuracy with respect to the target sentence and in terms of agreement with the noun in the recall attempt. The results are presented in Table 3. It can be seen that, in the areas of the grammar for which

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

the translation task revealed evidence of learning (albeit only for some subjects in the case of verbs), there were also correlations with Block 1 memory performance. There were also correlations involving Block 5 memory, but these tended to be smaller, apart from the case of modifier agreement.

One possible explanation for these correlations is that Block 1 memory reflects rapid learning of the rules relating to modifier and verb inflections, and those subjects who learned these rules in Block 1 were those who performed best in the translation task. However, two aspects of the data suggest that it was good memory for the forms themselves, rather than rapid learning of rules, that was responsible for the correlations. First, when modifiers were scored for agreement, the correlation involving Block 1 memory was not significant. If subjects were rapidly learning rules, they might have been able to use this knowledge to increase levels of agreement in their recall attempts, but this did not appear to be the case. Second, correlations between memory performance over the first four trials of the memory task and Block 1 translation-task performance were also computed, scoring modifiers for accuracy. The correlation for verbs was 0.623, $p < 0.01$, and that for modifiers was 0.657, $p < 0.01$. It seems unlikely that some subjects would have learned the relevant rules from just four sentences when other subjects apparently learned very little from 40.[7] Thus, there is evidence that learning of both form-form relations (noun-modifier agreement), and form-function relations (functions of verb inflections) was closely related to recall of the relevant forms as measured even very early in the memory task.

## Discussion

The results of Experiment 1 will now be discussed in relation to the issues raised in the introduction.

***Research Question 1: Memory as a Determinant of Inductive Learning.*** There was rather striking confirmation of the hypothesis that memory ability early in the experiment correlates with degree of learning as measured in the translation posttest. Indeed, learning correlated more highly with memory performance at the beginning of the experiment than at the end. It would appear from this result that there is at least some sense in which knowledge of grammatical rules emerges out of memory for input and that individual differences in memory ability that are apparent even in the earliest stages of exposure have consequences for ultimate levels of learning. These results suggest that the learning occurring in this experiment can be characterized as data-driven.

***Research Question 2: Memory for Input as a Consequence of Inductive Learning.*** Improvements in memory performance that could be attributed to learning of generalizations concerning the structure of the language were detected. The strongest evidence for this was found in verb inflections. Given that many subjects showed large gains in memory for verb inflections without

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

apparently having grasped their functions, it will be assumed that the memory gains were due to knowledge of the segmentability of (even novel) verbs into stems and one of four familiar suffixes—that is, the ability to chunk novel verbs in a morphologically appropriate way.[8] There was also evidence for improved noun-modifier agreement in recall at the end of the memory task, which in this case corresponded to superior performance in the translation task. Thus it appears that, as learning proceeds, memory for input is improved and there are cyclical interactions between learning and memory of the kind proposed by Ellis (1996).

Apart from the case of verbs, discussed above, there was no evidence of dissociations between performance in the memory and translation tasks. Indeed, the fact that translation task performance correlated with performance in Block 5 of the memory task for verbs and modifiers strongly suggests that the two tasks were sensitive to the same knowledge. Note in particular that in the case of modifiers this correlation increased when Block 5 recall was scored for agreement rather than accuracy. This suggests that subjects were able to use their knowledge of noun-modifier agreement to aid the regeneration of surface form and that this knowledge was consciously available to guide their choices in the translation task.

***Research Question 3: Induction of Functions.*** In general, learning in this experiment was confined to aspects of form. Subjects became sensitive to the segmentation of verbs into stems and affixes and to euphony between nouns and modifiers. This result might be regarded as support for the view that form-function relations cannot be learned incidentally during a task that is ostensibly focused on surface information. However, other aspects of the data suggest that subjects were sensitive to meaning during the memory task: (a) There was good learning for possessive type, (b) in post-memory-task interviews, many subjects reported that they paid particular attention to vocabulary, and (c) there was some evidence of function learning in the case of verbs, but only for subjects who displayed good memory for the verb endings in Block 1 of the memory task. A plausible explanation of (c) is that only subjects with good memory for verb endings at the start of the experiment came to an appreciation of the identity of the inflections rapidly enough to permit induction of their functions within the time scale of the experiment. Subjects with poorer memory for verb endings may have acquired the ability to chunk the verbs appropriately but were unable to induce functions. It is possible that, if the problem of identifying the inflections could be eased in some way, then even subjects with poor memory would be able to learn functions. Experiment 2 addressed this possibility.

Nouns were a clear exception to the relationship between memory and learning. Recall of noun inflections was very good from the beginning of the memory task, and yet there was very little evidence of learning their functions and no correlation between translation task performance and memory performance. However, there is evidence from the translation task that subjects ex-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

perienced considerable difficulty working out how the semantic dimensions of this paradigm related to the inflections. In the translation task, performance on verbs improved in response to feedback, whereas this was not the case for nouns (see Table 2). Additionally, examination of the 12 usable[9] think-aloud protocols collected in Block 1 of the translation task revealed that, although all subjects grasped the idea that verb inflections encoded the sentence subject, there was far more confusion over the function of the noun inflections. Here the choice was simply between the plural and singular forms; for example, *cavallo* "horse" versus *cavalli* "horses." Although number was indeed given as the basis for the choice on 61% of the occasions that a justification was provided, gender was given 34% of the time (gender is irrelevant in terms of the choice provided). Five subjects vacillated between justifications in terms of number and gender on different items.[10] Extrapolating to the memory task, one might conclude that, even under conditions in which subjects are including sufficient semantic information in their memory representations of sentences to induce certain form-function correspondences, certain mappings may not be learned because of the difficulty of identifying the appropriate functions. This suggests that knowledge of form-function mappings may not simply emerge out of memory representations of input material as assumed by the data-driven mode of learning.

Indeed, a conservative interpretation of the correlations between early memory and learning obtained in this experiment is that early memory ability was only important for learning the segmentation of verbs, nouns, and modifiers into stems and affixes, and that the involvement of memory in learning the rules relating to the distribution of these affixes remains an open issue. Subjects with good memory for input would have rapidly identified the relevant inflections and hence would have had more time to induce the rules governing their distribution. This would have resulted in a correlation between early memory and learning, but it would not mean that the process of learning distributional rules was data-driven.

One way of further exploring the relationship between memory and the induction of rules governing the distribution of morphemes within the time scale of a laboratory experiment is to increase the salience of the relevant morphemes in the input, thereby obviating the need to identify them through a prior segmentation process. This was the purpose of Experiment 2.

## EXPERIMENT 2

Three modifications were made to the procedure in order to increase the salience of the grammatical morphemes in the input sentences and to enable subjects to recognize the relevant grammatical morphemes without having to rely on memory-dependent segmentation processes. First, the grammatical morphemes were highlighted when presented visually; for example, *Am**ano le** donn**e** grech**e*** "They love the Italian women" (for previous studies of the effect of input enhancement that makes use of similar techniques for focusing

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

learner attention on form, see Doughty, 1991; Leeman, Arteagoitia, Fridman, & Doughty, 1995; Long & Robinson, 1997, and Robinson, 1997a, for reviews). Second, on recall trials, subjects were presented immediately with the highest level recall cue (stems of all content words), so that the task itself was reduced to recall of grammatical morphemes. Third, subjects were pretrained on the vocabulary used in the training sentences (Blocks 2–4) of the memory task. The verbs were taught in the infinitive form (e.g., *amare*, "to love"; *studiare* "to study"), the nouns in the singular form (e.g., *libro*, "book"; *donna* "woman"), and the adjectives in the masculine singular form (e.g., *greco*, "Greek"; *rotto*, "broken"). It was assumed that vocabulary pretraining would provide a stable basis in memory for one form of each word so that any variation in its ending would become more salient. Alternatively, following suggestions made by VanPatten (1990), subjects might be expected to find it easier to attend to grammatical morphemes if processing of content words for meaning is facilitated.

The procedural modifications outlined above would be expected to increase the degree of attention that subjects pay to grammatical morphemes in the input. However, given the theoretical framework outlined in the introduction, there are two ways in which this could influence learning. On the one hand, it could increase the degree of maintenance rehearsal of grammatical morphemes, thereby improving retention and leading to enhanced learning of relevant distributional rules. On the other hand, it could increase the extent of elaborative processing of grammatical morphemes, helping the subject to form appropriate hypotheses or to perform more specific tests on already-established hypotheses in memory. Thus, although a facilitative effect of input enhancement might be anticipated, the essential question is whether learning gains will be achieved through improvements in memory, thereby implicating data-driven learning processes. The two specific research questions addressed by Experiment 2 were therefore:

***Research Question 1: The Effect of Input Enhancement on Memory.*** Will memory performance be enhanced with respect to Experiment 1, especially as measured in the early stages of the memory task?

***Research Question 2: The Effect of Memory on Learning.*** Will early memory performance correlate with translation-task performance?

## Subjects

Sixteen subjects participated in this experiment, mean age 23 years. They were paid £10.00 for participating. Ten of them were Cambridge University undergraduates studying science subjects, two were postgraduate students in sciences, and one was a postgraduate studying history (whose first degree was in economics). Two were recent graduates working locally, both of whom had studied arts subjects (but not languages). Seven had taken French at General Certificate of Secondary Education level, and eight had taken both French

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

and German. Two subjects had some knowledge of Latin, and three subjects had very basic knowledge of Spanish. Twelve subjects said that they had used a foreign language outside a formal educational context, and this was an average of 3.75 years prior to the experiment.

## Method

***Materials and Design.*** Exactly the same presentation lists were used as in the real B condition of Experiment 1. The materials for the translation task were also identical.

***Procedures.*** The experiment now began with a vocabulary-learning phase. The computer displayed the English translations of the Italian words, grouped by grammatical category. They consisted of the four verbs (in the infinitive form), four adjectives (in the masculine singular form), and six nouns (in the singular form) from the training sentences to be used in Blocks 2, 3, and 4 of the memory task. By clicking on each word, the subjects could both see and hear its Italian equivalent (spoken by the same female native speaker of Italian who had recorded the sentences for the memory task). The subjects worked over the items at their own pace until they felt that they had memorized them. They then proceeded to a self-test phase in which the English words were presented in a random order one at a time, the subjects produced the Italian aloud, and they then clicked on the English word to check their response against the spoken and written Italian equivalent. After completing this test they could exit from the program if they were 100% correct or else start again. The experimenter monitored the subjects' oral responses from an adjacent room to ensure that they were indeed 100% accurate on the final self-test (ignoring minor pronunciation and stress errors). It took the subjects between 5 and 10 minutes to learn the 14 words.

The changes to the memory task were as follows: The written form of each Italian sentence was now altered so that the grammatical morphemes were in red, and the remainder of the sentence was in black (e.g., *Am**iamo la** donn**a** grec**a***, "We love the Greek woman"; *Ogni weekend ripar**o la loro** macchin**a***, "Every weekend I repair their car"; bold indicates red). On recall trials, the meaning representation and its context were first presented alone and the subjects were required to produce the English equivalent aloud. On clicking "continue," they were given the stems of the noun, verb, and, when relevant, the adjective just as for the final help level in the previous version. It should be noted that this procedure speeded up the recall trials so that, whereas the interval between presentation and recall for the first sentence in each pair was the same as before, in the case of the second sentence it was rather shorter. It was in order to increase the recall interval in the present procedure that subjects were now required to reinterpret the semantic figures in English at the start of each recall trial.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 4.** Block 1 translation performance (%) in Experiment 2

| Condition | Verb Function | Noun Function | Article-Noun Agr. | Adjective-Noun Agr. | Possessive-Noun Agr. | Possessive Type |
|---|---|---|---|---|---|---|
| Exp 2 | 73 | 81 | 53 | 87 | 52 | 78 |
| Diff | +25* | +18* | +20* | +15* | –8 | –13 |

*Note.* Exp 2 = Experiment 2, Diff = Experiment 2 minus Experiment 1, *$p < 0.05$, Agr. = agreement.

## Results

***Memory Task.*** Memory-task performance was scored as in Experiment 1. Levels of memory performance were remarkably similar to those in the real condition of Experiment 1. In Block 1, overall percent correct recall averaged over the five form types was 59%, compared to 57% in Experiment 1. In Block 5 the difference between the two experiments was slightly greater, being 82% in Experiment 2 and 74% in Experiment 1, but the difference was not significant, $F < 1.0$ in an analysis of variance on the Block 5 data alone, and there was no interaction with form type, $F < 1.0$. In order to assess the impact of vocabulary pretraining on early memory, the data from Block 2 were also analyzed. Here too there was no significant difference in overall level of recall, which was 69% in Experiment 2 and 65% in Experiment 1.

***Translation Task.*** The translation data were scored as in Experiment 1. Table 4 shows the percent correct for each form type as well as the difference with respect to the scores in the real condition of Experiment 1. Only the Block 1 data are reported because this provides the best indication of the relative levels of learning in the memory task itself.

An analysis of variance was conducted to compare translation performance in the real condition of Experiment 1 with that in Experiment 2. Experiment was a between-subjects factor and form type was a within-subjects factor. The main effect of experiment was not significant, $F(1, 30) = 3.12$, $p < 0.1$. However, there was a significant interaction between experiment and form type, $F(5, 150) = 4.36$, $p < 0.001$. As can be seen in Table 4, not all form types showed the same degree of improvement with respect to Experiment 1. Two-tailed $t$ tests were performed to evaluate the difference between experiments for each form type, and the results are indicated in Table 4. Significant differences were obtained for verbs, nouns, articles, and adjectives. The only form not to show improved learning in Experiment 2 was possessive-noun agreement (and possessive type, but this was nearly at ceiling in Experiment 1). By and large, the greatest improvements were obtained for those areas of the grammar that showed the least evidence of learning in Experiment 1.

***Relationships between Memory and Translation Task Performance.*** As in Experiment 1, there was a high degree of variability between subjects in memory and translation task performance. For Block 1 memory, $M = 54\%$, $SD = 19.1$, range 32–95. For Block 5 memory, $M = 79\%$, $SD = 15.2$, range = 35–95. For Block

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 5.** Correlations between translation and memory performance in Experiment 2

| Translation Task Performance | Block 1 Memory | Block 5 Memory |
|---|---|---|
| Modifier-noun agreement | .395 (.594*) | .079 (.232) |
| Article-noun agreement | .092 (.159) | .722** (.676**) |
| Noun function | –.192 | .049 |
| Verb function | .481 | .677** |
| (Overall) | .774*** | .619** |

*Note.* *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Correlations when memory performance was scored for agreement are shown in parentheses.

1 translation, $M = 70\%$, $SD = 16.7$, range = 42–97. Correlations were calculated between Block 1 and Block 5 memory and Block 1 translation. The results are presented in Table 5.

The results of the correlational analyses contrast with those of Experiment 1. When considered on a form-type by form-type basis there were now no significant correlations between Block 1 memory and Block 1 translation. The independence of learning and early memory is seen most clearly in the case of articles, for which the correlation between translation task performance and Block 1 memory approached zero. Indeed, an analysis of memory for articles through Blocks 2–4 of the memory task revealed no significant correlations with translation-task performance, whether recall was scored in terms of accuracy ($r = .33, .29, .29$ for Blocks 2, 3, and 4, respectively) or agreement ($r = .31, .26, .18$). For verbs, memory-task performance began to show correlations with the translation task by Block 3 ($r = .39, .72, .73$ for Blocks 2, 3, and 4, respectively).

Modifiers provide the only evidence for a relationship between early memory and learning, but only when scored for agreement. It appears that even in Block 1 some subjects were sensitive to the euphony between nouns and modifiers. This conclusion is reinforced by the fact that, in Block 1, 17% of incorrectly inflected modifiers agreed with the incorrectly inflected noun (compared with none in Experiment 1). Subjects who exhibited high levels of agreement in Block 1 went on to produce high levels of noun-modifier agreement in the translation task. However, in contrast to Experiment 1, recall of modifiers in Block 5 did not correlate with translation task performance. This is presumably because many subjects became better at recalling modifier inflections, but only those who were sensitive to agreement in Block 1 showed evidence of having learned the agreement rule.

Finally, and rather surprisingly, despite the lack of correlations between Block 1 memory and Block 1 translation for individual form types, highly significant correlations were obtained when overall scores were considered. What this suggests is that subjects' learning was broadly constrained by their Block 1 memory performance, an issue that will be explored further below.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

## Discussion

The introduction of highlighting and vocabulary pretraining in Experiment 2 radically altered the pattern of results compared to Experiment 1. Significant improvements in translation-task performance were obtained for verbs and nouns, which indicates that most subjects were now able to learn the functions of the relevant grammatical morphemes. There was also an improvement in learning the form-form relations involved in article-noun and adjective-noun agreement. In general terms, these improvements would seem to demonstrate the effectiveness of devices that draw the subjects' attention to specific aspects of the input through highlighting.[11] This effect was much more robust than in previous attempts to measure the impact of highlighting on learning. For example, Robinson (1997a) examined the effect of highlighting on learning a phonological constraint on dative alternation. In the transfer test, the critical measure of rule learning was the rejection rate for ungrammatical sentences (this required the subject to realize when a novel verb was incorrectly used in a double-object construction). Rejection rates were numerically, but not significantly, higher for subjects who had been trained in the enhanced condition (a meaning-oriented task plus highlighting of relevant aspects of form) than in the incidental condition (meaning-oriented task with no highlighting). The enhanced condition was significantly better than an implicit condition (involving judgments about the relative positions of words), but the difference in basic task means that this comparison does not provide a measure of the impact of highlighting as such. Alanen (1995) examined inductive learning of three locative suffixes in a semiartificial form of Finnish, again under meaning-oriented task conditions. For the subjects in the enhanced condition the suffixes were italicized, whereas for the subjects in the control condition they were not. In a subsequent production task it was found that the control-group subjects were far more likely than the enhanced group to produce bare, uninflected forms. However, in terms of accuracy the two groups were both very poor when compared with other groups of subjects who had received explicit instruction. The subjects in the enhanced group produced a large number of nontargetlike suffixes, indicating that they knew that some kind of suffix was required but did not have a clear idea of the appropriate forms or functions. In the present case, the use of highlighting in conjunction with a form-oriented memory task seems to have had a much greater impact on learning of both forms and functions.

Let us turn now to the specific research questions.

***Research Question 1: The Effect of Input Enhancement on Memory.*** Learning gains with respect to Experiment 1 were achieved in the absence of improvements in memory performance, especially as measured in Blocks 1 and 2 of the memory task. Highlighting and vocabulary pretraining did not appear to improve learning outcomes via their impact upon early memory.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

***Research Question 2: The Effect of Memory on Learning.*** As the correlational analyses showed, with the exception of modifiers, learning of individual form types was now largely unrelated to their recallability early in the memory task. Assuming that the present experiment is primarily sensitive to the process of inducing the rules governing the distribution of morphemes (highlighting having obviated the need to segment bound morphemes from the input), then it would appear that subjects' eventual knowledge of such rules was not dependent upon their level of memory for the relevant morphemes early in the experiment.

The only area of the grammar to show any hint of a relationship between Block 1 memory and learning was modifiers, but only when scored for agreement with the noun. As argued above, this can be interpreted as showing that some subjects learned the modifier-noun agreement rule very rapidly and began to use it to aid the recall process even within Block 1, a pattern that was not obtained in Experiment 1.[12] Given the simplicity of this rule (it does not even depend upon knowing all of the possible inflections because it can simply be stated as a copying rule), rapid learning under conditions of highlighting in some subjects is not surprising. What is more surprising, however, is that subjects whose recall attempts exhibited low levels of noun-modifier agreement over the first 8 sentences of Block 1 exhibited relatively poor levels of learning in the translation task despite receiving 32 more sentences. Given that even subjects with poor initial memory might have been expected to learn a simple rule within the time scale of the experiment, this result provides further evidence that learning distributional rules is subject to factors other than mere memory for input.

Learning the distributional properties of morphemes appears not to depend upon early memory for the relevant forms—a result that might be taken as indicating a conceptually driven mode of learning. By implication, in Experiment 1, the correlation between early memory and translation performance for verbs and modifiers must have been due to the importance of early memory for learning segmentation, thereby indicating a data-driven mode of learning in this domain.

However, if learning in Experiment 2 was essentially conceptually driven, why were there correlations between early memory and learning when overall scores were considered? One possible explanation of this is that Block 1 memory performance provides an indication of an individual's general ability to encode novel forms in phonological long-term memory (PLTM) and that whatever learning mechanism was operating in Experiment 2 still drew on this general ability. This argument would be bolstered if Block 1 translation were also found to correlate with some independent measure of PLTM. The vocabulary-learning phase of the present experiment could, in principle, provide an indication of subjects' PLTM ability because it could indicate how readily subjects form stable long-term phonological representations of Italian words. Unfortunately, however, the procedure used in the present experiment did not permit

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

an accurate assessment of how quickly the vocabulary was learned; subjects were free to browse through the items as much as they wished before trying the self-test phase. In Experiment 3, a modified procedure was used in order to test whether vocabulary-learning efficiency was correlated with inductive learning. A post hoc analysis of the present data using the relatively crude measure available will be reported along with that analysis.

## EXPERIMENT 3

Apart from introducing a more refined measure of vocabulary learning, this experiment included modifications designed to test whether learning in Experiment 2 was conceptually driven. If such learning is construed as an active process of hypothesis formation and testing, then learning outcomes should be enhanced by the provision of feedback on recall trials of the memory task. General frameworks for modeling conceptually driven inductive learning lay particular emphasis upon the revision of rules in response to negative feedback (Anderson, 1983; Holland et al., 1986). In the specific context of second language acquisition, Klein (1986) has proposed that learning occurs when a mismatch is perceived between the learner's own output and that of others. There are also a number of empirical demonstrations of the effectiveness of feedback under conditions that might be regarded as conducive to conceptually driven learning (Carroll & Swain, 1993; Tomassello & Herron, 1988).

In Experiment 3, when subjects made incorrect responses on recall trials in Blocks 2–4, they were provided with the correct answer and were required to change their response accordingly. Because subjects who make more recall errors receive more feedback under the present procedure, learning outcomes should be particularly enhanced in subjects with poor memory performance in Block 1, thereby suppressing correlations between overall Block 1 memory and overall translation-task performance.[13]

### Subjects

Sixteen subjects participated in the experiment, mean age 26 years (the mean was 24 years excluding one subject who was 53 years old).[14] They were paid £10.00 for participating. Six of them were Cambridge University undergraduates, seven were postgraduates, two were postdoctoral researchers, and one was a part-time lecturer. Subject areas were mainly scientific apart from two studying law, one economics, one developmental studies, one management studies, and one music. At the General Certificate of Secondary Education level, 10 subjects had taken only French, 1 had taken only German, and 4 had taken French and German. Other languages that the subjects had learned to a basic level at some point in their lives were Russian (three subjects), Spanish (one subject), Latin (two subjects), and Chinese (two subjects). One subject was an English-Bengali bilingual. Eight subjects said that they had used a foreign language outside a formal educational context, and this was an average of 5.75 years prior to the experiment.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

## Method

***Materials and Design.*** These were identical to Experiment 2.

***Procedures.*** For the main experiment, the only changes to the procedure concerned the recall trials of the memory task. Subjects were presented with the same recall cues as in Experiment 2, except that the display also contained text entry prompts at the end of the verb stem, in the article position, the possessive-adjective position (if relevant), the end of the noun stem, and at the end of the adjective stem (if relevant). Responses could only be entered in the correct order, and the elements of the noun phrase were therefore arranged in the correct order on the screen. After reaching the end of the sentence, subjects were able to correct their responses if they wished.

Feedback was provided only during Blocks 2–4 of the memory task. This was to enable direct comparison with previous experiments with respect to Blocks 1 and 5. During feedback trials, whenever the subjects entered an erroneous response the correct morpheme immediately appeared either above or below it for 2 seconds. The subjects' incorrect response then disappeared and they were required to try again before proceeding to the next field. Thus, feedback was provided incrementally, on a morpheme-by-morpheme basis, and not at the end of the sentence.

The vocabulary-learning phase was now administered in such a way as to permit a more sensitive measurement of learning time than had been possible in Experiment 2. The whole procedure now followed the format of the self-test phase of the previous version. English words appeared one at a time and the subjects attempted to produce the Italian. They then clicked to see and hear the correct response. Subjects cycled through the list of 14 words in a fixed order (grouped by grammatical category) and, when they felt more confident, they switched to a randomized presentation. They continued cycling through the items until they could produce the Italian for each item (barring minor errors of stress and pronunciation).

## Results

***Memory Task.*** Memory-task performance in Blocks 1 and 5 was scored as in Experiment 1. There were no significant differences in level of recall in Blocks 1 and 5 for any form type with respect to Experiment 2. Given that in previous experiments there was a strong relationship between overall performance in Block 1 of the memory task and overall Block 1 translation-task performance, it is important to note that overall Block 1 recall was no different from that in Experiment 2 (55% and 60% in Experiments 2 and 3, respectively, $F < 1.0$).

***Translation Task.*** The translation data were scored as in Experiment 2. Overall performance in Block 1 was slightly worse than in Experiment 2, at 64% as opposed to 71% in Experiment 2. However, an analysis of variance in which experiment was a between-subjects factor and form type a within-sub-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 6.** Correlations between translation and memory performance in Experiment 3

| Translation Task Performance | Block 1 Memory | Block 5 Memory |
|---|---|---|
| Modifier-noun agreement | .526* (.593*) | .462 (.570*) |
| Article-noun agreement | .150 (.333) | .735** (.744***) |
| Noun function | .042 | .006 |
| Verb function | .011 | .456 |
| Overall | .670** | .697** |

*Note.* *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Correlations when memory performance was scored for agreement are shown in parentheses.

jects factor showed that this difference was not significant, $F < 1.0$, nor was there any interaction between experiment and form type, $F < 1.0$. Individual two-tailed $t$ tests showed that only in the case of nouns was there a significant 15% decrement in performance with respect to Experiment 2, $t(30) = 2.21$, $p < 0.05$.

***Relationships between Memory and Translation Task Performance.*** Correlations between Block 1 memory performance and Block 1 translation task performance were calculated as in Experiment 2. These are presented in Table 6.

The pattern of correlations is broadly similar to that obtained in Experiment 2. Verb functions and noun-article agreement failed to show any correlation between Block 1 memory and Block 1 translation, whereas for article-noun agreement, at least, the correlation involving Block 5 memory was again very strong. Once again, modifiers showed evidence of rapid sensitivity to agreement in Block 1. On this occasion this was also evident in Block 5 agreement (possibly because translation-task performance was more variable than in Experiment 2).

***Correlations with Vocabulary Learning Efficiency in Experiments 2 and 3.*** In this experiment, the efficiency with which the subjects acquired the vocabulary was measured in terms of the number of times they had to cycle through the vocabulary items in either fixed or random order before being able to produce the Italian for at least 12 out of the 14 words (barring minor errors of pronunciation or stress). The mean score was 4.94 cycles, $SD = 1.625$, range 2–8 cycles. This measure was highly correlated with overall Block 1 memory performance, $r = .732$, $p < 0.01$. The correlations between vocabulary learning efficiency and Block 1 translation task scores are shown in Table 7, along with equivalent correlations performed on the data from Experiment 2 (in which vocabulary learning was measured in terms of the cruder measure of time taken to achieve at least 14 out of 16 on a self-test and data was only available for 14 out of 16 subjects).

From these analyses, a very clear pattern emerged that was replicated across the two experiments: Vocabulary learning efficiency correlated with

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

**Table 7.** Correlations between vocabulary learning efficiency and Block 1 translation task performance in Experiments 2 and 3

| Translation Task Performance | Experiment 2 Vocabulary | Experiment 3 Vocabulary |
|---|---|---|
| Modifier-noun agreement | –.537* | –.781*** |
| Article-noun agreement | –.707** | –.628** |
| Noun function | .177 | –.189 |
| Verb function | –.393 | –.327 |
| (Overall) | –.570* | –.766*** |

*Note.* $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

learning of article-noun and modifier-noun agreement but not with learning the functions of the verb and noun inflections.

## Discussion

Feedback did not improve learning with respect to Experiment 2. In fact, in one area of the grammar—namely, noun functions—translation-task performance was actually significantly worse.[15] This result is not consistent with a conceptually driven view of learning. If learning were due to processes of hypothesis testing and revision, then feedback should have had an effect. This is not because learning of the present grammar cannot, in principle, be improved by feedback in the form of a correct answer: An effect of feedback was obtained in the translation task of Experiment 1 for most areas of the present grammar. Of course, there may be other reasons why feedback was ineffective: Perhaps subjects did not perceive the recall trials as providing a learning opportunity (although why this should be so if they were operating in a conceptually driven learning mode is not clear), or the feedback may have been insufficiently coordinated with their own learning processes (although according to the notion of learning by "noticing the gap," Schmidt & Frota, 1986, feedback should have directed their learning processes).

Additional evidence that a simple conceptually driven mode of learning should not be assumed too readily comes from the correlational analyses in which, as in Experiment 2, overall memory performance in Block 1 correlated with overall translation-task performance. A highly significant correlation was also obtained between overall translation-task performance and vocabulary-learning time. If learning were purely conceptually driven, it is not clear why measures that can be assumed to tap phonological long-term memory are related so closely to learning. The picture is further complicated by the fact that correlations between vocabulary-learning time and translation-task performance were only obtained for rules that concern elements of surface form (article-noun agreement and modifier-noun agreement) and not for form-function mappings. In a sense, this is surprising because one might have thought that,

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

because learning vocabulary involves learning associations between new forms and meanings, the strongest correlations should have involved learning functions of grammatical morphemes.

An attempt to integrate these findings into a general characterization of learning in the present experiments will be presented below.

## GENERAL DISCUSSION

In considering the results of the present experiments, it is useful to distinguish two aspects of learning: first, learning to segment words into stems and affixes, and, second, learning the rules that govern the distribution of grammatical morphemes.

### Learning Segmentation

On the basis of the results for verbs in Experiment 1, it would appear that learning segmentation and function are dissociable. Many subjects appeared to become sensitive to the segmentation of novel verbs into unfamiliar stems and familiar affixes while having very little idea of the specific functions that those affixes encoded. The correlations between very early memory performance and learning were taken as suggesting that subjects with good memories were able to rapidly segment out the bound morphemes, which gave them more opportunity to learn the distributional properties of those morphemes within the confines of the experiment. This would suggest that learning segmentation is a memory-dependent, and essentially data-driven, process.[16]

Peters (1983) describes a number of learning procedures for segmenting the speech stream in first language acquisition. The most relevant in the present case is the "Match1" strategy: "If two units share any phonologically similar portion, the shared portion can be segmented out and stored as a unit, and so can the residues" (Peters, 1983, p. 37). This strategy can be conceived as operating either "on-line" or "off-line." An example of off-line learning would be if the learner has *ascoltiamo* "we hear" and *ascolto* "I hear" stored as units, or "instances" (to use a term from Holland et al., 1986) in memory and through the operation of the Match1 strategy abstracts the units *ascolt-*, *-iamo*, and *-o*. In on-line learning, the learner only needs *ascoltiamo* stored as an instance in memory. On hearing *ascolto*, they are able to segment out the shared portion (*ascolt*) and store the residues separately (*-iamo* and *-o*). In both cases, however, it is obvious that the accuracy with which input forms are encoded in memory is a critical determinant of learning. A simple verbatim memory task such as that used here would provide a fairly direct indication of the quality of the representations of instances that the learner is able to form.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

## Learning Distributional Rules

The complex pattern of results obtained in Experiments 2 and 3 does not fit cleanly with either a data-driven or conceptually driven conception of learning. On the one hand, there were no form-specific correlations between learning and early memory performance, and highlighting and vocabulary pretraining enhanced learning without affecting early memory. These findings might be taken to imply a conceptually driven mode of learning. On the other hand, Experiment 3 showed no effect of feedback, and in both Experiments 2 and 3 it was not the case that early memory was totally irrelevant to learning because correlations were obtained between Block 1 memory and Block 1 translation when overall scores were considered.

Further insights into the nature of the learning process are provided by subjects' comments in the debriefing administered after the memory task in Experiment 2. When asked whether they were actively trying to work out the grammar, some of the responses were as follows: "Well I think one *recognizes* the patterns after a while" (emphasis in original); "I could see there was a pattern coming"; "You eventually pick up on the endings" (two subjects); "No. It's just something I noticed. I mean when I started off I was doing it just by sounds, and then I obviously started noticing patterns like *-iamo* is coming up for 'our' and *-iano* for 'their' which obviously helped things along"; "About halfway through I started seeing patterns, and then I started to remember the endings." The use of verbs such as *recognize*, *pick up*, and *see* suggests a relatively passive mode of learning. The final comment also illustrates how memory for endings was regarded as a consequence of learning rather than a cause. First, subjects "started seeing patterns" and then they "started to remember the endings." One may wonder, however, how patterns were detected if endings were not reliably recalled; yet this is precisely what the lack of a correlation between early memory and learning suggests was happening.

One way of integrating the present findings is to assume a distinction between simple verbatim memory for grammatical morphemes in sentences and the construction of instances that can support learning of their distributional properties. On every trial of the memory task, the subjects will have engaged in maintenance rehearsal of the input sentence (this was forced on them by the procedure) that will have secured a certain degree of long-term memory encoding of the component morphemes as a simple linear string of items. Such an encoding would have been sufficient for knowledge of segmentation to emerge in Experiment 1 because, as indicated above, the relevant instances for learning were defined at the level of the single word. However, for knowledge of the distributional properties of morphemes to be learned, it may have been necessary for the subjects to focus attention on specific relations between forms in the current input sentence (e.g., the cooccurrence of *la* and *-a* in ***la*** *donn****a***, "the woman") or between a form and an aspect of the sentence meaning (e.g., *-iamo* and WE in *am****iamo la*** *donn****a*** *grec****a***, "We love the Greek woman"); that is, to construct an instance for learning in which specific as-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

sociations were prominent. This process cannot be characterized as either maintenance rehearsal (simply cycling phonological material around the phonological loop) or elaborative processing (relating current input to long-term memory schemata). Rather, it might be thought of as integrative processing. Graf (1994) defines integrative processing as follows:

> Integration focuses on connections among the units that define an individual item . . . ; these kinds of connections are formed or strengthened when the subject either perceives coherence among separate stimulus components (e.g., under the guidance of preexisting representations or gestalt laws like proximity or common fate) or conceives a structure for processing target features as a single entity. (p. 685)

Given the pace of presentation of material in the present experiment, and the continual demands of the verbatim recall task, subjects probably found it difficult to engage in integrative processing on every trial. As one subject put it when asked if he was looking for patterns, "I was trying to, but I didn't find there was enough time in between to sort of recognize and remember one of the patterns, that's all." It seems likely that they could only perform this kind of analysis on a minority of trials and probably not until they had become relatively familiar with other more general aspects of the language (e.g., word order and lexis). Thus, even though in principle integrative processing should have enhanced recall, the relative rarity of this happening early in the experiment would have made any correlations between early memory and learning unlikely (although noun-modifier agreement does seem to have been an area in which integrative processing occurred very early in Experiments 2 and 3). According to this analysis, learning in Experiments 2 and 3 was essentially data-driven, but the requirement for instances to be appropriately structured meant that only a restricted sample of the input entered into the learning process.

If learning distributional rules is critically dependent upon the subjects initially paying attention to relations between elements in the input, then it follows that even the simplest rules might not be learned if the subjects for some reason fail to attend to those relationships. This might explain why, in Experiments 2 and 3, some subjects apparently learned the very simple rule of noun-modifier agreement from the first 8 sentences, whereas other subjects failed to do so despite receiving a further 32 sentences. Some subjects simply failed to attend to the appropriate relations, whereas others did so from the very beginning of the experiment. In general, however, the enhanced learning in Experiment 2 with respect to Experiment 1 could be interpreted as showing that highlighting and vocabulary pretraining encourage appropriate integrative processing.

### Intention to Learn and Awareness of Product

The traditional dichotomy between implicit and explicit learning contrasts learning with and without both intention to learn and awareness of product.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

In the present case, there was evidence that subjects had high levels of awareness of the product of learning, but in what sense was the learning process intentional?

In fact, it is not difficult to envisage a learning mechanism that leads to conscious knowledge without an intention to learn. For example, according to the Match1 strategy, aspects of the current input "resonate" with memory traces of previous input sentences that are similar to it along some dimension (see Hintzman, 1986, for an intuitively appealing model of how such resonances occur). What learners are conscious of is the structuring effect of these memory processes upon their perception of the current sentence; they need not be conscious of the long-term memory representations that lead to the structuring effect (see Baars, 1988, for discussion of how unconscious "contexts" shape perception). The memory comparison processes that lead to segmentation can be implicit, but the results of those processes are explicit. As Ellis (1996) argues in relation to the acquisition of the phonotactics of the L2:

> Although learners need not be aware of the processes of such pattern extraction, they will later be aware of the product of these processes because the next time they experience that pattern it is the patterned chunk that they are aware of in working memory, not the individual components. (p. 109)

The learning of agreement rules and functions could follow the same principle. For example, the learner might pay attention to the co-occurrence of *la* and the noun inflection *-a* in the phrase ***la*** *donn****a*** "the woman," and this may be stored in long-term memory as an instance. This knowledge structure could then interact with the occurrence of ***la*** *macchin****a*** "the car" in a subsequent sentence resulting in the formation of the general template ***la*** NOUN + ***a***. At the same time, the phrase *la macchina* would be perceived as a specific instantiation of this template; an illustration of how, in Graf's (1994) terms, preexisting representations might cause the subject to perceive coherence among separate elements of the input. Repetitions of this perception of coherence in further examples would strengthen the template and increase the probability that it could be utilized in the translation task.

According to the above analysis, interactions between short- and long-term memory are automatic and do not require conscious recall of prior instances. In this sense, learning can occur without intention. However, this is strictly the case only for learning segmentation. For learning distributional rules, the learner must actively impose structure on the input material, and this implies a certain degree of intention. Presumably, highlighting facilitates learning because it invites the learner to actively look for patterns relating elements of the input. However, in the context of a memory task such as that employed here, it is not clear whether there is an intention to learn the grammar of the target language or whether the intention to see structure is part of the general goal of achieving accurate recall. It is possible that learning grammar is in fact

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

achieved incidentally because the type of integrative processing that improves recall also happens to be that which is appropriate for learning grammar. This would be consistent with the failure to obtain an effect of feedback in Experiment 3: If the subjects had an intention to learn grammar, then feedback should have been of benefit; if they were merely attempting to integrate elements of individual sentences as a means of improving recall, then any extra information after recall had occurred would be irrelevant.

## PLTM and Learning Distributional Rules

If learning distributional rules was essentially data-driven, based on the encoding of (appropriately structured) instances in memory, then learning should still have been constrained by general PLTM encoding ability, even if there were no form-specific correlations between early memory and learning. This is precisely the pattern obtained in Experiments 2 and 3. The fact that correlations were strong for form-form relations is to be expected given that the demands on PLTM are highest in this case (two forms have to be integrated in the same instance). The absence of correlations for form-function relations suggests that there were factors other than PLTM that influenced either the degree to which subjects paid attention to meaning during the memory task (which would determine the probability of constructing instances in which forms were related to meanings) or the ease with which they were able to identify the relevant functions (which would determine the probability of constructing appropriate instances).

Similar observations of a close relationship between vocabulary and grammar learning have been made in research on learning the forms and functions of novel grammatical morphemes by normal children and children with Specific Language Impairment. Swisher and Snow (1994) and Nakamura, Plante, and Swisher (1990) found that children's ability to learn the function of a novel bound morpheme was related to their ability to learn the novel vocabulary items that carried that inflection within the learning phase of the experiment. However, Swisher and Snow (1994) found that good vocabulary learning was only a necessary, but not sufficient, condition for being able to generalize the morpheme appropriately to new words—there were many subjects who, despite high vocabulary scores, failed to generalize the morpheme. This is to be expected if, as argued here, learning segmentation is a prerequisite to learning functions, but the induction of functions is subject to factors other than phonological long-term memory ability.

More generally, Bates and Goodman (1997) and Bates, Bretherton, and Snyder (1988) have stressed the relationship between grammar and vocabulary in both first language acquisition, abnormal language development, and aphasia. For example, in one study (Bates et al., 1988, Study 6) there was a highly significant correlation between a measure of vocabulary development at 20 months and grammatical development at 28 months. Superficially, this result appears to be compatible with the present proposal that grammar learning is

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

dependent upon long-term phonological memory ability. However, Bates and Goodman presented data from a variety of other studies that showed a highly consistent relationship between rate of development of vocabulary and grammar. They suggested that grammatical development is dependent upon vocabulary size and offered reasons for this ranging from "perceptual bootstrapping" (the more words you know the easier it is to identify and pay attention to grammatical morphemes in the input) to the central role that, in certain linguistic theories, the lexicon might play in encoding grammatical knowledge. From this viewpoint, phonological memory ability can only be related to grammatical development via its influence upon the rate of vocabulary acquisition (for which there is good evidence, Gathercole & Baddeley, 1993), but it does not influence grammatical development directly. In contrast, the present experiments suggest that, when adult subjects are provided limited exposure to novel language input and when their vocabulary size is held constant, their ability to induce certain types of grammatical rules is directly related to their general phonological memory ability. It has been proposed that this is because PLTM is required to encode the phonological components of instances that then drive the inductive learning process.

## Implications

What are the implications of the present results for language learning outside the laboratory? First and foremost, they highlight the importance of PLTM (specifically the ability to form accurate long-term memory representations of novel phonological forms) as a source of individual differences in learning grammatical rules by induction. This effect was not confined to acquiring the ability to segment grammatical morphemes from the input but also to learning those rules governing the distribution of grammatical morphemes that could be described at the level of form. A surprising aspect of the results was how wide the variation in PLTM was between subjects who, in other respects, might be regarded as rather similar in general intellectual ability and how closely variation in PLTM was related to learning outcomes. The extent to which this close connection between memory and inductive learning is a product of the use of a memory task as the vehicle for presenting input, or whether it generalizes to inductive learning under other task conditions, remains to be seen.

Second, the interpretation of the present results has emphasized the importance of integrative processing and attention in forming appropriately structured instances for learning distributional properties of morphemes. Learners do not simply absorb input and abstract the underlying statistical regularities as easily as seems to be implied by the assumption that inductive learning is simply the process by which structure "emerges" from memories of past input (Decoo, 1996) and that is appealed to by a "natural" approach to teaching (Krashen, 1981, 1985). Even rules that might be thought of as simple and transparent (e.g., plural marking of nouns) might fail to be induced if the

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

relevant forms and meanings are not segmented out of the input stream and intentionally associated together in an instance. This perspective lends support to pedagogic approaches that seek to facilitate learning by manipulating the way in which learners attend to, or process, the input (Doughty & Williams, 1997; Long & Robinson, 1997; Loschky & Bley-Vroman, 1993; Terrell, 1991; VanPatten, 1996; White, Spada, Lightbown, & Ranta, 1991).[17,18]

Finally, it has been argued that inductive learning does not have to be seen as either an explicit process of hypothesis formation and revision or as an implicit process of unconscious rule abstraction. Rather, a middle position is possible whereby knowledge that is sufficiently explicit to transfer to other tasks can result from the implicit structuring effect of long-term memories of (appropriately structured) instances on the learner's conscious perception of current input without the learner necessarily having an explicit intention to learn grammatical rules as such. Inasmuch as the vast majority of the participants in the present study found this mode of learning to be more enjoyable and more engaging than traditional forms of explicit instruction that they had experienced in the past, it seems worth considering to what extent simple memory tasks may be used to promote inductive learning in a way that is less taxing than traditional rule discovery.

(*Received 18 February 1998*)

## NOTES

1. The present memory task is not meant to be an analogue of any real-world language learning activity (if it were perceived as such it would be likely to encourage conceptually driven as opposed to data-driven learning). Like many tasks used in experimental psychology it is simply designed in such a way as to reveal something about the functioning of certain basic psychological processes. The extent to which those processes might also be operative in other situations outside the laboratory is a separate issue (see Implications, General Discussion).

2. One may wonder whether these subjects are representative of the average language learner. Although they are likely to be of above average intelligence, the restriction of the sample to science students was intended to limit their degree of linguistic sophistication. In general, however, concerns about the representativeness of the sample are mainly relevant to the absolute level of learning achieved, which is a function not only of the ability of the learner but also of the complexity of the grammar. Because the present concern is with the relative learnability of different aspects of the grammar under different task conditions, the issue of generalizability to different populations of learners is not seen as a major problem. A working assumption is that similar outcomes could be obtained with other types of learners by manipulating the complexity of the grammar.

3. As one reviewer pointed out, this procedure might draw the subjects' attention to the importance of word endings. This means that even the limited degree of learning that occurred in this experiment could overestimate that which might occur when no clues as to the potential importance of word endings can be gleaned from the task. At the same time it cannot be claimed that the results obtained in this experiment were entirely due to the fact that subjects' attention was drawn to grammatical morphology because in Experiments 2 and 3, in which attentional allocation was manipulated more overtly, a qualitatively different pattern of results was obtained.

4. Most previous research on learning miniature languages has employed grammaticality judgments as a measure of learning. The grammaticality judgment task was not adopted here because it would have made it difficult to evaluate knowledge of form-function mappings and because violations in concord might have been easily missed. It also seemed possible that, by mixing items in which concord is violated with items in which it is not, subjects might enter a hypothesis testing mode during the grammaticality judgment test itself.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

5. These between-condition comparisons may inflate the effect of agreement because of a possible reduction in sensitivity to agreement in the corrupted Italian condition.

6. The remaining four subjects in the real condition scored 75% or above in the translation task for verbs. The present subgroup analysis is therefore important in order to establish that the memory gain for verb inflections in the memory task is not primarily due to these subjects.

7. An alternative possibility is that there was in fact very little actual learning of noun-modifier agreement but that performance reflected either some subject's prior knowledge of similar languages or a general disposition in some subjects to provide nouns and modifiers with the same ending. Such prior dispositions could have been suppressed through exposure to corrupted Italian, leading to differences in translation- and memory-task performance between the two groups. This could also account for the correlations between Block 1 memory and learning in the real Italian group because those subjects with a prior disposition to provide nouns and modifiers with the same ending would have shown superior memory performance (even early in the experiment) as well as superior translation-task performance. However, as already noted, this would not explain the reduction in the correlation when Block 1 memory is scored for agreement (on this view it should increase) nor the significant improvement in adjective-noun agreement between Blocks 1 and 5 of the memory task (see Table 1). In general, the only way to control for the effect of prior knowledge in this experiment would have been either to administer a version of the translation task as a pretest (which would be clearly inappropriate because of its sensitizing effect) or to administer the translation task to a separate group of subjects who did not perform the memory task (although given that the effect of prior knowledge or dispositions is obviously not homogeneous in the population—as shown by the correlational analyses—it is doubtful that this would constitute a sufficiently reliable control without testing far greater numbers of subjects).

8. Without other tests of subjects' segmentation ability, this must remain an assumption, although it is one that seems justified on the basis of the wealth of evidence for the effect of chunking on memory (Ellis, 1996).

9. Four subjects did not provide enough justification for any of their decisions to make analysis worthwhile.

10. Interestingly, mention of number was more common when the referent was plural (19 mentions of number vs. 3 mentions of gender) than when it was singular (6 mentions of number vs. 11 mentions of gender). This suggests that, although plurality was a salient feature of the semantic representation of the English sentence, singularity was not, and that in such cases the subjects considered other semantic dimensions that could plausibly be encoded morphologically, namely gender. As one reviewer pointed out, this recourse to gender is rather strange given that all of the subjects were native speakers of English, although it is also true that all of them had knowledge of either French or German.

11. In the case of nouns, it is also possible that, because they had learned each noun in its singular form, encountering the same noun with a different ending drew subjects' attention not only to the change in form but also to the change in meaning (i.e., number). Although in principle this same contrast was provided across occurrences of nouns in Experiment 1, the vocabulary-learning phase may have allowed subjects to establish more stable representations of both forms and meanings than was possible in the sentence memory task. Thus, the improvement in learning noun functions may not simply be a product of the increased salience of the noun ending as such, but also to the increased salience of the relevant aspects of the sentence meaning.

12. In contrast, in Experiment 1 the correlation between Block 1 memory and learning for modifiers was greater when scored in terms of accuracy and there were no cases in which a modifier was forced to agree with a misrecalled noun ending. This pattern is more consistent with the view that accurate recall of modifier endings was a prerequisite to learning agreement because, as argued earlier, it facilitates recognition of the endings as inflections.

13. It might be thought that the present feedback procedure would facilitate data-driven learning as well as conceptually driven learning because it provides greater amounts of positive as well as negative evidence. However, this is to assume that subjects make an effort to encode the correct answer in memory to the same extent that they attempt to encode the original input sentence. There is no reason why they should do this in the context of the task because after the recall trial there is no requirement to remember the sentence.

14. This subject achieved one of the highest levels of learning in the experiment.

15. A possible explanation of this is that on recall trials in which the article was corrected, the correct article may have then served as a recall cue for the remainder of the noun phrase. Such a tendency, by focusing on the form-level association between a noun and its accompanying article, would have diverted attention from the association between noun endings and functions.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

16. It should be noted that in first language acquisition research the standard view of, say, verb inflection learning is that the child first establishes connections between unanalyzed forms (e.g., *walks*, *walked*) and the meanings they express (*walk* + present + third-person singular, *walk* + past + third-person singular), and only later segments off the relevant inflections to enter into each cell of the paradigm; that is, segmentation occurs only after analysis of meaning (Pinker, 1984; see also Terrell, 1991, for a similar proposal in relation to second language acquisition). However, it must be noted that, in the present case, the use of a limited set of stems in the training sentences and the absence of any irregular forms meant that the task of segmentation was relatively simple compared to naturalistic contexts, and this may have led to the situation in which segmentation was achieved in the absence of knowledge of function.

17. Of course, it may be objected that this emphasis upon the way the learner intentionally structures the input is exaggerated because of the highly limited exposure provided in the present experiments, and that outside of the laboratory learning is better characterized as the appropriation of the rules of the input language through very gradual, "nonselective" (Berry & Broadbent, 1988), and implicit learning mechanisms. However, as noted in the introduction, controlled laboratory studies have not as yet produced compelling evidence for implicit learning of abstract rules. Thus, until we have better evidence that abstract, generalizable, and productive knowledge of grammar can indeed emerge implicitly and nonselectively from memories of past input, there seems no reason to assume that the kinds of learning mechanisms described here do not operate in more naturalistic contexts, only that given the varied and uncontrolled nature of naturally occurring input and the varying purposes the learner has, the probability of encoding appropriately structured instances in memory is much reduced compared to what can be achieved under more controlled conditions.

18. The present conclusions only relate to learning morphological paradigms and those agreement phenomena that are evident at the level of surface form. Whether memory and attention play any lesser role in the acquisition of other aspects of grammar, and in particular core properties of grammar that are theoretically regarded as falling in the domain of UG, is an interesting empirical question.

## REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.

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

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

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

Baddeley, A. (1990). *Human memory: Theory and practice*. London: Erlbaum.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

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*, 507–586.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251–272.

Bialystok, E., & Ryan, E. (1985). A metacognitive framework for the development of second language skills. In D. L. Forrest-Pressley, G. MacKinnon, & T. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 207–252). New York: Academic Press.

Bley-Vroman, R. & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 245–261). Hillsdale, NJ: Erlbaum.

Carroll, J. B., & Sapon, S. M. (1959). *The Modern Language Aptitude Test*. New York: The Psychological Corporation.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback. *Studies in Second Language Acquisition, 15,* 357–386.

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.

Davis, S. (1992). *Connectionism: Theory and practice*. Oxford: Oxford University Press.

Decoo, W. (1996). The induction-deduction opposition: Ambiguities and complexities of the didactic reality. *IRAL, 34*, 95–118.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–276.
DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379–410.
Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of second language relativization. *Studies in Second Language Acquisition, 13*, 431–469.
Doughty, C., & Williams, J. (1997). Pedagogic choices in focus on form. In C. Doughty and J. Williams (Eds.), *Focus on form in classroom SLA* (pp. 197–261). New York: Cambridge University Press.
Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5*, 289–318.
Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). New York: Academic Press.
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., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition, 19*, 145–171.
Fischer, R. A. (1979). The inductive-deductive controversy revisited. *Modern Language Journal, 63*, 98–105.
Gathercole, S., & Baddeley, A. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
Graf, P. (1994). Explicit and implicit memory: A decade of research. *Attention and Performance, 15,* 681–696.
Hammerly, H. (1975). The deduction/induction controversy. *Modern Language Journal, 75*, 15–18.
Hintzman, D. L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.
Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
Klein, W. (1986). *Second Language Acquisition*. New York: Cambridge University Press.
Knowlton, B. J., & Squire, L. S. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 79–91.
Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
Krashen, S. (1985). *The Input Hypothesis: Issues and implications*. London: Longman.
Leeman, J., Arteagoitia, I., Fridman, B., & Doughty, C. (1995). Integrating attention to form with meaning: Focus on form in content based Spanish instruction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.
Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing and instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1022–1050.
Long, M. H., & Robinson, P. (1997). Focus on form: Theory, research, practice. In C. Doughty and J. Williams (Eds.), *Focus on form in classroom SLA* (pp. 15–41). New York: Cambridge University Press.
Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning* (pp. 123–167). Clevedon, UK: Multilingual Matters.
MacWhinney, B. (1982). Basic syntactic processes. In S. A. Kuczaj (Ed.), *Language development: Vol. 1, Syntax & semantics* (pp. 73–136). Hillsdale, NJ: Erlbaum.
Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237.
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.
Nakamura, M., Plante, E., & Swisher, L. (1990). Predictors of novel inflection learning by pre-school age boys. *Journal of Speech and Hearing Research, 33*, 747–754.
Perruchet, P. (1994). Learning from complex rule-governed environments: On the proper functions of non-conscious and conscious processes. *Attention and Performance, 15,* 811–835.
Peters, A. M. (1983). *The units of language acquisition*. New York: Cambridge University Press.
Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.
Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 88–94.

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

Reber, A. S., Walkenfield, F., & Hernstadt, F. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888–896.
Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45,* 283–331.
Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18*, 27–67.
Robinson, P. (1997a). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–247.
Robinson, P. (1997b). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45–99.
Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.
Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.
Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21–42). Oxford: Oxford University Press.
Swisher, L., & Snow, D. (1994). Learning and generalization components of morphological acquisition by children with specific language impairment: Is there a functional relation? *Journal of Speech and Hearing Research, 37*, 1406–1413.
Terrell, T. C. (1991). The role of grammar instruction in a communicative approach. *The Modern Language Journal, 75*, 52–63.
Tomassello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics, 9*, 237–246.
VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition, 12*, 287–301.
VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.
White, L., Spada, N., Lightbown, P., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12*, 416–432.
Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 115–145). New York: Academic Press.

# APPENDIX

## DETAILS OF SENTENCE COMPOSITION AND A LISTING OF MEMORY TASK ITEMS

### Real Italian

*Real A training sentences* ($n = 24$). Half of the sentences contained an adjective and half a possessive adjective. Words were selected as described below:

Verbs: There were four different verbs. Each verb stem occurred six times, and bore each inflection at least once, and two inflections twice. Across all verbs each inflection occurred six times.

Adjectives: There were four different adjectives. Each occurred three times, and not more than once bearing the same inflection. Across all adjectives, each inflection occurred three times.

Possessive adjectives: Each possessive type occurred three times. For "my," "his," and "ours," the possessive adjectives only occurred bearing two of the possible inflections, one of which was borne twice. *Loro* "their" occurred with nouns bearing two of

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

the possible endings and with one of them twice. The remaining eight cells of the paradigm were reserved for test items (see below).

Nouns and articles: There were 24 different nouns. There were equal numbers (i.e., six) of masculine singular, masculine plural, feminine singular, and feminine plural nouns. Thus there were six occurrences of each type of article.

*Real B training sentences* (*n* = 24). These were constructed as above except that this time only six different nouns were used. Each noun occurred twice in its singular and plural form. In addition, the contexts were changed so that the English translation would contain the definite article.

*Test sentences* (*two sets of n* = 8). For each set of test sentences, half contained an adjective and half contained a possessive adjective. Unlike for the training sentences there was no repetition of verbs and adjectives. None of the verbs, adjectives, or nouns had occurred in the training sentences. The composition of each set of eight test sentences was as follows:

Verbs: There were eight different verbs, two bearing each inflection.

Adjectives: There were four different adjectives, one bearing each inflection.

Possessive adjectives: There was one possessive of each type ("my," "his," "ours," "theirs"). Four of the remaining cells of the complete paradigm were used for each set of test items.

Nouns and articles: There were eight different nouns. Each inflection occurred twice, as did the accompanying article.

*Translation sentences* (*three sets of eight sentences each*). Note that, although some of the words in the translation sentences were the same as in the test sentences used in the memorization phase, familiarity effects were minimized in the first (critical) block of translation trials by ensuring that the relevant items had only been encountered in Block 1 of the memorization task.

Verbs: The verbs in Blocks 1 and 2 had the same stems as verbs used in the test sentences, but different forms were required. Each inflection occurred twice. For Block 3, eight new verbs were used, each inflection occurring twice.

Adjectives: Four unique adjectives were used in each set of sentences. Each inflection occurred once.

Possessive adjectives: For the items in Blocks 1 and 2 the same forms of the possessive adjectives were used as in the test sentences.

Nouns and articles: Eight different nouns were used in each block. Each inflection occurred twice, as did the accompanying article.

#### Corrupted Italian

Each of the real A Italian training or test sentences had a corrupted Italian counterpart that contained the same verb, noun, and adjective stems, and was paired with the same meaning representation. However, although overall each grammatical morpheme occurred with the same frequency as in the real Italian, they were redistributed over the sentences in a completely non-rule-governed fashion.

*Training items.* Verbs: Each verb stem now only occurred with one inflection unique to that verb (e.g., *riparo* "I repair" occurred six times). This meant that each verb + in-

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

flection now expressed each possible function ("I," "he" or "she," "we," "they") at least once, and two functions twice.

Adjectives: Each adjective stem now only occurred with one unique inflection (e.g., *rotto* "broken" occurred three times). Adjectives never occurred with an accompanying noun bearing the same ending (e.g., *tetti rotto* "broken roof," *donne famosa* "famous woman").

Possessives: Only four forms of the possessive adjectives were used: *nostro*, now used to express "my," *loro* "his," *miei* "ours," *sue* "theirs." They never occurred with a noun bearing the same ending.

Nouns and articles: Noun inflections were redistributed so that each ending expressed a plural referent as often as a singular one. Each article now occurred with a noun carrying each inflection at least once, and for two inflections, twice.

Corrupted versions of the real Italian practice sentences were also written.

*Test items.* Note that corrupted Italian test items were only presented in Block 1. However, because of the rotation of sets of test items around Blocks 1 and 5, it was necessary to form corrupted Italian versions for both sets of real Italian test items.

Verbs: Exactly the same forms were used as in the real Italian condition. However, the contexts were changed so that each function was now encoded by two different inflections, neither of which was the correct one.

Adjectives: Each inflection occurred once, but it was never the same as that carried by the noun.

Possessives: The same possessive forms were used as in the corrupted Italian training sentences. They were paired with the same meanings. There was never any agreement between the possessive adjective and the noun.

Nouns and articles: Each article occurred twice, but with a noun carrying a different inflection on each occasion (and never with the correct inflection). Noun endings were equally frequent, but there was no consistent correlation with number.

#### Items Used in the Memory Task of Experiment 1 (Real B) and Experiments 2 and 3

The sentences were presented in the following order except for the fact that for half of the subjects the Block 1 and Block 5 sentences were exchanged.

Block 1 (Test):

| | |
|---|---|
| *Odio la musica moderna* | "I hate the modern music" |
| *Spesso ascoltiamo i loro dischi* | "Often we listen to their records" |
| *Ogni weekend lavano le nostre finestre* | "Every weekend they wash our windows" |
| *Insegna i ragazzi poveri* | "He teaches the poor children" |
| *Ammirano il teatro russo* | "They admire the Russian theatre" |
| *Di solito porto la mia valigia* | "Usually I carry my suitcase" |
| *Ogni anno cambia il suo lavoro* | "Every year he changes his job" |
| *Di solito piantiamo le piante insolite* | "Usually we plant the unusual plants" |

Block 2 (Training):

| | |
|---|---|
| *Amo il libro greco* | "I love the Greek book" |
| *Spesso ripariamo le sue macchine* | "Often we repair his cars" |
| *Spesso usa le mie grammatiche* | "Often he uses my grammars" |
| *Amiamo la donna greca* | "We love the Greek woman" |

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press

| | |
|---|---|
| *Studia i vini famosi* | "She studies the famous wines" |
| *Ogni weekend riparo la loro macchina* | "Every weekend I repair their car" |
| *Spesso usano il nostro tavolo* | "Often they use our table" |
| *Ogni weekend studiano la grammatica italiana* | "Every weekend they study the Italian grammar" |

Block 3 (Training):

| | |
|---|---|
| *Ripara i tavoli rotti* | "He repairs the broken tables" |
| *Spesso usiamo il loro vino* | "Often we use their wine" |
| *Ogni sera studio i miei libri* | "Every evening I study my books" |
| *Amano la donna famosa* | "They love the famous woman" |
| *Ogni weekend riparano le macchine rotte* | "Every weekend they repair the broken cars" |
| *Di solito uso i miei tavoli* | "Usually I use my tables" |
| *Studiamo la nostra grammatica* | "We study our grammar" |
| *Amo le donne greche* | "I love the Greek women" |

Block 4 (Training):

| | |
|---|---|
| *Studiano i vini italiani* | "They study the Italian wines" |
| *Usiamo la loro macchina* | "We use their car" |
| *Studia il nostro libro* | "She studies our book" |
| *Ama le donne famose* | "He loves the famous women" |
| *Ripara il tavolo rotto* | "He repairs the broken table" |
| *Uso le sue grammatiche* | "I use his grammars" |
| *Ripariamo i suoi libri* | "We repair his books" |
| *Amano il vino italiano* | "They love the Italian wine" |

Block 5 (Test):

| | |
|---|---|
| *Curano i bambini malati* | "They look after the sick babies" |
| *Ogni weekend visita la sua casa* | "Every weekend he visits her house" |
| *Spesso fumiamo le loro sigarette* | "Often we smoke their cigarettes" |
| *Vendo il gelato delizioso* | "I sell the delicious ice cream" |
| *Spesso compriamo la poesia romantica* | "Often we buy the romantic poetry" |
| *Spesso guardano i nostri vestiti* | "Often they look at our clothes" |
| *Di solito risparmio il mio denaro* | "Usually I save my money" |
| *Spesso ruba le collane costose* | "Often she steals the expensive necklaces" |

https://doi.org/10.1017/S0272263199001011 Published online by Cambridge University Press