SSLA, **27**, 269–304. Printed in the United States of America.
DOI: 10.1017/S0272263105050138

LEARNING WITHOUT AWARENESS

John N. Williams
University of Cambridge

Two experiments examined the learning of form-meaning connections under conditions where the relevant forms were noticed but the critical aspects of meaning were not. Miniature noun class systems were employed, and the participants were told that the choice of determiner in noun phrases depended on whether the object was "near" or "far" from the subject of the sentence. What they were not told was that the choice of determiner also depended on the animacy of the noun. Most participants remained unaware of this correlation during the training and test tasks; yet when faced with a choice between two determiners for a noun, they chose the one that was appropriate to the noun's animacy at significantly above-chance levels, even though that combination had never been encountered during training. This ability to generalize provided evidence of learning form-meaning connections without awareness. In both experiments, there was a correlation between generalization test performance and knowledge of languages that encode grammatical gender. This points to the importance of prior knowledge in implicit learning.

Implicit learning occurs without intention to learn and without awareness of what has been learned. It is clearly of great educational importance to know what can and cannot be learned in this way and what factors might make some individuals more successful implicit learners than others. At a theoretical level, the study of implicit learning can help us come to understand the nature of unconscious learning mechanisms, their relationship to other cognitive constructs such as memory and attention, and their interactions with existing

Many thanks to Ronald Leow, Daphnée Simard, and the anonymous *SSLA* reviewers for their helpful comments on an earlier draft of this article.

Address for correspondence: John N. Williams, Research Centre for English and Applied Linguistics, University of Cambridge, 9 West Road, Cambridge, CB3 9DP, UK; e-mail: jnw12@cam.ac.uk.

knowledge in the mind of the learner. The experiments reported here address these issues in the context of learning form-meaning connections.

To clarify the basic concepts used in this article, it is useful to consider a specific model of the relevant cognitive functions. Consider the *embedded processes* model of working memory proposed by Cowan (1999), which has also been used by Robinson (1995) as a framework for discussing memory and attention in an applied linguistic context. The rudiments of this model are straightforward, and yet they capture insights from a number of other approaches to memory and attention. External stimuli, or internally generated associations, activate memory representations. There is no known limit on the number of representations that can be activated in memory, but they will only remain active for a brief time. Focal attention acts to boost the activation level of representations, allowing them to remain active over extended periods. In the case of external stimuli, attention also increases the detail of encoding, or the quality of the representation. Within "global workspace" models of consciousness (e.g., Dehaene & Naccache, 2001), the function of attention is to make information available to a wider range of processes than would otherwise be the case. For example, attention allows information to be manipulated strategically to accomplish a task and to control intentional behavior. However, only a limited amount of information can benefit from attentional enhancement at any one time. Which representations are selected for focal attention is partially under voluntary control, but selection also occurs involuntarily through the attentional orienting system (which, for example, automatically directs attention to unexpected stimuli). Both Cowan and Dehaene and Naccache assumed that the contents of focal attention and conscious awareness are coextensive. "I assume that, in neurally intact individuals, the information in the focus of attention is the same information that the person is aware of, which is also the same information to which a central capacity limit applies" (Cowan, 1999, p. 89). If a person can verbally report or intentionally respond on the basis of some information, then that individual is assumed to have attended to or been aware of that information. If a person cannot, then that individual is assumed not to have attended to it or been aware of it.

Cowan's (1999) model integrated the notions of attention and memory, but what about learning? Tomlin and Villa (1994) used the term *detection* to refer to activation of information in memory and argued, in common with Cowan, that memory representations can be activated even without the individual being aware of them (in Cowan's terms, without them being attended to). Convincing evidence has been provided in support of this position (Merikle, Smilek, & Eastwood, 2001). Tomlin and Villa went slightly further, however, and suggested that representations that are activated, but not amplified by focal attention, can contribute to learning. This is a contentious issue, and it lies at the core of the research reported in this article. Cowan pointed out that, in most situations, what gets activated are novel combinations of memory elements, whether they are novel combinations of features making up a novel object or novel combinations of familiar objects. The issue for learning, then, is whether associations between those elements can be learned even when those elements do not individually receive focal attention. Can people learn associations between stimuli that they are not aware of? Schmidt (1990, 1994, 2001) has argued against such a view, claiming essentially that there is no learning without focal attention to the relevant stimuli. Stimuli that do not receive focal attention and that are outside awareness might activate preexisting memory representations subliminally. However, new memory representations cannot be formed unless the relevant stimuli enter into conscious awareness or, in Schmidt's terms, are "noticed."

In practical terms, the simplest way to determine whether a stimulus has been noticed is to rely on verbal report. Leow (1997, 2000) pioneered this approach to studying the relationship between awareness and learning. His studies examined learning examples of irregular (stem-changing) verbs in Spanish, verbs that in the third-person singular and plural forms of the preterit tense undergo a vowel change in the stem (e.g., *repetir* becomes *repitió*). Learners were exposed to these forms in the context of crossword puzzle tasks, and they were instructed to think aloud as they worked through the clues and wrote in the solutions. In the study by Leow (2000), the crossword puzzle was constructed so that the stems were already provided and learners had to fill in only the endings. Nevertheless, half of the participants noticed the peculiar stem change in the critical items; that is, they commented upon it in their verbal report. When tested on the same verbs in recognition and production posttests, they showed large gains in accuracy as compared to a pretest. The verbal reports from the remaining participants showed no evidence of having noticed the stem change in the verbs, and there was no significant difference between their pretest and posttest scores. Assuming that the critical aspects of the stimuli were processed sufficiently to be detected, this study demonstrated that there is no learning in the absence of noticing, a conclusion consistent with studies in experimental psychology that showed no learning of new information without focal attention (Jiménez & Méndez, 1999; Logan & Etherton, 1994).

However, we must be careful in defining learning. Leow's (1997, 2000) experiments concerned item learning: The issue was how well participants could remember the forms that they had encountered during the training task. However, in the case of language learning, productivity is a central concern as well: Can people learn generalizations across items and apply them to new situations? As Schmidt (2001) pointed out, noticing only determines whether specific instances of linguistic input are encoded in memory. Once this has occurred, generalizations across instances can, in principle, be formed with or without the learner becoming aware of what those generalizations are or how they were arrived at (see also Robinson, 1995). At one extreme, there is entirely explicit learning involving conscious comparisons between current and previous instances of input and the formation and testing of hypotheses. At the other extreme, there is entirely implicit learning; the learner has no awareness of either the process or product of learning.

In cases where a learner becomes aware of generalizations, one can speak of "awareness at the level of understanding" (Schmidt, 1990, p. 145). Leow (1997) found that this level of awareness could also be identified in the verbal protocols of his participants. There are in fact two rules governing stem changes in Spanish: an *e* changes to an *i* (e.g., *mentir* – *mintió*) and an *o* changes to a *u* (*dormir* – *durmió*). Participants who verbalized at least one of these rules during the crossword puzzle task showed better posttest performance than those who merely commented on the stem change in individual items; that is, awareness at the level of understanding produced more learning than awareness at the level of noticing. Presumably, participants who grasped the rule were less reliant on rote memory for individual items in the posttest. Leow did not test for generalization to new verbs, but in a different domain, Rosa and Leow (2004) showed that awareness at the level of understanding was associated with greater gains from pretest to posttest even when new items were used in the test material. This is, of course, to be expected. What is interesting is that in this experiment, awareness at the level of noticing also produced significant pretest to posttest gains, suggesting that some implicit rule learning might indeed have occurred. On the other hand, participants who did not report noticing showed no learning. Noticing is a necessary condition for learning, but understanding might not be. Again, this is consistent with work in experimental psychology that has demonstrated that people can learn something about the abstract structure of artificial (finite state) grammars under conditions where they are attending to the relevant material but are unable to describe the relevant rules (Knowlton & Squire, 1996; Mathews et al., 1989; Reber, 1976). One could debate exactly what is meant by *abstract* in such studies and how relevant the rule systems are to natural language, but it does seem to be possible to generalize beyond individual instances without awareness of the underlying rules.

All the work cited so far has concerned learning at the level of form. It is not implausible that humans (and, indeed, some primates) are equipped with powerful statistical learning mechanisms for tracking associations between forms in the input (Hauser, Newport, & Aslin, 2001; Saffran, Newport, & Aslin, 1996) and that the brain encodes this information in such a way as to permit extraction of generalizations. But how do these ideas apply to the learning of form-meaning connections?

AWARENESS AND LEARNING FORM-MEANING CONNECTIONS

If it is the case that learning depends on awareness at the level of noticing the relevant stimuli, then in the case of learning form-meaning connections, the implication is that it is necessary to attend to both form and meaning. The first time a word is encountered, the learner might only be able to guess at its meaning or encode a rough approximation with the help of a teacher or dictionary. This would constitute a working hypothesis as to the meaning of the word.

What is critical here is that not only are form and meaning attended to, but that they are also integrated by a form-meaning connection. In all but the simplest cases, the true meaning of the word would need to be derived by abstracting over a number of form-meaning pairings. This could be achieved by either implicit or explicit means. In the explicit case, when the word is encountered again, the learner might explicitly recall the previously hypothesized meaning. If this does not seem to fit the current context, the hypothesis could be modified. Through a cyclic process of explicit hypothesis formation and testing, the learner might come to acquire the meaning of the word. However, it is also possible that the true meaning of a word is abstracted through implicit learning mechanisms. This process might be no more than a consequence of the way in which instances of form-meaning connections are stored in memory, as proposed by connectionist memory models (McClelland & Rumelhart, 1985).

Many researchers of both first (L1) and second (L2) language acquisition stress the role of attention in learning form-meaning connections. In L1 acquisition, Merriman (1999) argued that the child must attend to a form in relation to a certain aspect of meaning. The task for the researcher is to determine the factors that influence the child's distribution of attention, such as the shape bias (Smith, 1999) or the search for distinctive features (Merriman). For example, when faced by a second label for an object, the principle of *mutual exclusivity* is violated and the child is assumed to "seek out, discover, and dwell on features that are uniquely associated with the second label" (Merriman, p. 350). This sounds very much like an explicit process of hypothesis formation and testing.

With regard to L2 acquisition, Doughty and Williams (1998) stressed not only attention to form and meaning but also their integration within the learning episode. It is not enough that the relevant form and meaning are attended to; the learner must actually hypothesize a connection between them. This is an elaboration of the concept of noticing within the context of learning form-meaning connections. N. Ellis (1994) went further and argued that learning form-meaning connections is dependent on declarative memory—the same memory systems that are responsible for the storage of personally experienced events, broadly referred to as *episodic memory*. Remembering that you think that the word *ravenous* means something like "very hungry" is therefore like remembering that you ate a particularly large breakfast last Sunday morning, and working out the precise meaning of *ravenous* involves explicit recall of contexts of use of the word.

N. Ellis' (1994) arguments are based on the fact that variables that are known to affect episodic memory also influence vocabulary learning (e.g., intelligence, depth of processing, and metacognitive and mediational strategies). Gupta and Dell (1999) made a similar argument on purely computational grounds, pointing out that the arbitrariness of form-meaning mappings makes them very difficult to assimilate using the kinds of implicit (in their view, connectionist) learning mechanism that underlie skill learning. Important evidence is also provided by research on amnesia. For example, amnesics, such as H. M. (Gabrieli, Cohen, & Corkin, 1988) or SS (Verfaellie, Koseff, & Alexander, 2000), can learn novel word forms, as assessed by tests of implicit memory such as fragment completion and priming. However, they seem to have great difficulty learning new vocabulary; for example, they are unable to recognize or use words that they have been exposed to over long periods after the onset of amnesia. Amnesics are unable to recall specific experiences; they have impaired episodic memory. Their inability to acquire vocabulary thus implies that episodic memory is involved in learning form-meaning mappings.

However, more recent work suggests that amnesics can learn vocabulary after all. Vargha-Khadem et al. (1997) examined three young adults who had suffered brain damage in infancy and become diagnosed as amnesics during their school years. Amazingly, their academic progress is reported to be in the normal range. Despite having great difficulty in recalling a story after a short delay, recounting a day's events, or remembering when their lessons were, they showed normal vocabulary and reading development and general knowledge. Other studies have also provided evidence that even severe adult amnesics can acquire vocabulary and factual information. Kitchener, Hodges, and McCarthy (1998) reported on a patient who could not remember highly salient personal events (e.g., that his daughter had left home or that his mother had died), and yet he was able to distinguish real and fictitious public events and to choose the correct definitions of words that came into the language only after the onset of his amnesia. Verfaellie et al. (2000) described a severely amnesic patient, PS, who showed preserved semantic learning, including learning of novel vocabulary. These studies count against the claim that vocabulary learning is necessarily dependent on memory for specific episodes of word use.[1] Rather, the process of abstracting the meaning of words from experience occurs implicitly. Word meanings can be learned without explicitly recalling episodes of use. However, the amnesic literature does not shed any light on the importance of attention to form and meaning at encoding, which might still be necessary for learning form-meaning mappings.

The consensus, therefore, appears to be that learning form-meaning mappings requires attention to form and meaning and probably their integration at encoding; that is, there needs to be awareness at the level of noticing of a form-meaning connection. However, there is less certainty over whether explicit, episodic memory is necessarily involved in working out word meanings from instances of use or in demonstrating knowledge of word meanings in comprehension and production (i.e., whether there needs to be awareness at the level of understanding).

PREVIOUS EXPERIMENTAL RESEARCH

Very little implicit learning research has specifically examined form-meaning connections. A notable exception is a study by DeKeyser (1995) that employed a miniature artificial language with rich inflectional morphology for marking biological gender, number, and case. For example, if the sentence subject was plural, the verb carried the suffix *-it*, and if the subject was singular, the verb carried the suffix *-at*. The training task was sentence-picture matching; participants had to indicate whether a given sentence correctly described a picture. Crucially, mismatching pairs always involved vocabulary differences and never purely grammatical errors, so the training task did not encourage participants to attend to the critical form-meaning relationships. Learning was measured by a production task in which participants were presented with a picture and had to type the correct sentence in the artificial language. Performance was at chance in the critical test of generalization to new sentences, despite the fact that the participants had endured 20 learning sessions of 25 minutes each. Thus, there was no evidence of implicit learning even of a simple subject-verb number agreement rule, even though that rule was very similar to one in the participants' L1, English. Note that participants showed good test performance for forms that had occurred during training, so they had obviously been paying sufficient attention to encode the relevant forms in memory during the training task. However, the training task did not require them to attend to the relevant aspect of meaning nor its potential relationship to forms. The results therefore support the idea that attention to form and meaning is necessary for learning.

One can imagine a situation in which both the relevant form and the relevant meaning are attended to, but the participant does not integrate them at encoding. In this case, the form and meaning can be said to be noticed, but the connection between them is not. Such a situation is clearly difficult to engineer, but in a previous work (Williams, 2004, Experiment 1), I reported a study that attempted to do just this. The system comprised eight novel nouns learned as translation equivalents of common English words (e.g., *johombe/i* "monkey/s", *nawase/i* "vase/s") and eight articles that were learned as translation equivalents of English "the" (*ig*, *i*, *ga*, *ge*), "a" (*ul*, *ula*), and "some" (*tei*, *tegge*). The participants were not informed why there were four words for "the" and two each for "a" and "some." During training, the participants were presented with noun phrases formed by combining nouns and articles (e.g., *i johombi*, *ga nawase*, *ul johombe*, *ge nawasi*). For each phrase, they had to (a) repeat it aloud (to facilitate memory encoding), (b) indicate whether the noun referred to a living or nonliving thing, and (c) translate the phrase into English. The distribution of the articles was in fact governed by noun definiteness, number, and animacy (animates took *ig*, *i*, *ul*, *tei*; inanimates took *ga*, *ge*, *ula*, *tegge*). The training task explicitly drew the participants' attention to the animacy of the noun without hinting that this was relevant to the choice of article. During training, each noun appeared with only three of the possible four determiners licensed by the grammar, the fourth being withheld for testing generalization ability. The training period consisted of 15 repetitions of each noun phrase, each repetition occurring in a separate block of trials. This was followed by a test phase in which the participants had to translate English phrases (e.g., "the monkey"). They were presented with a choice between two possible noun phrases (e.g., *ig johombe / ga johombe*) where the articles were of the correct number and definiteness, but one violated the animacy rule. The critical test items concerned the noun phrases that had been withheld from the training phase, referred to as *generalization* test items. There were also items that had occurred during training, referred to as *trained* items. After performing this test, the participants were interviewed to ascertain their level of awareness of the system and how they had approached the task.

Somewhat surprisingly, even after the test phase, 30 out of the 37 participants were unaware that animacy was relevant to determiner usage—that is, they had no awareness at the level of understanding. However, their responses to the generalization test items were 61% correct, which was significantly above the chance level of 50%. It was also found that performance was correlated with the participants' language background. For the participants who spoke an L1 that encodes grammatical gender (referred to here as a *gender language*), generalization test performance was significantly better than for the participants whose L1 did not encode gender. Because there was also a correlation with overall number of gender languages known, $r = 0.572$, $p < .001$, it is not clear whether it was knowing a gender language as L1 or knowing many gender languages in general that was important. There was also a correlation between generalization test performance and phonological short-term memory (PSTM), $r = .50$, $p < .01$. However, given that PSTM was also correlated with number of gender languages, $r = .519$, $p < .01$, it is not possible to say whether the relevant underlying factor was memory ability or linguistic knowledge.

The previous experiment suggested that it is possible, in principle, to implicitly learn form-meaning connections when they are attended—but not integrated—at encoding. This is consistent with work in experimental psychology that has looked at learning associations under conditions where the stimuli are attended to but the participants are not aware of the relationship between them (Jiménez & Méndez, 1999; Logan & Etherton, 1994). The results count against the suggestion that learning form-meaning connections necessarily involves intentionally integrating form and meaning at encoding (Doughty & Williams, 1998). However, the correlation with knowledge of gender languages suggests that the situation might be more complex than this and that interactions with prior linguistic knowledge might also be important.

In Experiment 2 of Williams (2004), I attempted to reduce the potential impact of prior knowledge of gender languages by employing a system that was superficially less similar to article systems in languages that the participants were likely to know. The participants first learned that the novel words *gi* and *ro* meant "near" and the words *ul* and *ne* meant "far" (colors were used to distinguish the alternative forms for each meaning). It was then explained that these words functioned like the English definite article but also expressed how far away the object was. The participants were then provided with phrases such as *ul dog* and were told to construct sentence contexts that highlighted the distance relationship implied by the novel word (e.g., *I threw a ball to ul dog that was at the bottom of the garden*). As in the first experiment, there was in fact a correlation between determiners and noun animacy. Animate objects took *gi* and *ul* and inanimate objects took *ro* and *ne*. To force attention to the animacy distinction, the participants were instructed to construct different kinds of context for living and nonliving things. Living things were to be described in outdoor contexts, and nonliving things were to be described in indoor contexts. Thus, for *ro telephone*, an appropriate sentence would be "I was woken by ro telephone by my bed." To help the participants focus on the relevant constraints during sentence production, they used icons on the computer screen to represent the noun (a picture of the object), near or far (using geometric figures), and indoor versus outdoor (a picture of a house or a tree). There were six nouns per class. Two in each class occurred with both possible determiners during training, the remaining four only occurred with one determiner, and the withheld items were reserved for testing generalization ability.

Participants performed this task for between 128 and 160 trials, depending on time constraints. There was then a test phase in which figurative representations of training and generalization items were presented (e.g., a picture of a lion, the far symbol, and a tree) along with a choice between two noun phrases (e.g., *ul lion* vs. *ne lion*). The participants were instructed to choose the phrase that seemed more familiar or "better." After completing this task, only 3 of 17 participants tested expressed any awareness of the relevance of animacy (or the inside/outside distinction). Performance on generalization test items for the remaining 14 was not significantly different from chance (51.5%), whereas performance on test items that corresponded to trained items was highly accurate (81%). There were no correlations with language background factors. Thus, in contrast to Experiment 1 of Williams (2004), there was no evidence of implicit learning of form-meaning connections, even though the relevant forms and meanings were attended during the training task.

There are two potential problems with this experiment, however. First, the training task was rather arduous for the participants. It is known that—at least in some cases—implicit learning is reduced under dual-task conditions (for discussion in an applied linguistic context, see Carr & Curran, 1994). Therefore, overall task demands might have been an important factor. Second, the target system was rather small. There were only six nouns per category, and only four of these occurred with both possible determiners during training. It has been argued that a critical mass of exemplars might need to be encoded before generalizations can emerge (Marchman & Bates, 1994; Robinson & Mervis, 1998). The first experiment reported in the following section was therefore designed to address these problems.

A second problem is that the training tasks were highly artificial because of the requirement to explicitly draw the participants' attention to the relevant meaning feature, but not its relationship to the determiners. In the present experiments, a more natural situation was constructed in which participants merely listened to sentences containing the determiner-noun combinations (rather than generating them), indicated whether the novel word meant near or far, and formed an image of the general situation described by the sentence in anticipation of a test of memory for overall understanding of the sentence (not exact wording). Their attention was not explicitly drawn to the animacy feature at all; rather, this was implicit in the representation of the context. This task is less stressful and demanding than the sentence generation task. However, it also takes us one step further away from the noticing hypothesis. It seems unlikely that participants would spontaneously notice whether the relevant noun was animate or inanimate in this task. The task only requires them to notice the identity of the novel determiner. Any evidence of implicit learning of the form-meaning connection under conditions where only the form is attended would run counter to the prevailing emphasis on attention to form and meaning—and their integration—in discussions of learning form-meaning connections in the L1 and L2 learning literature.

An additional feature of the present experiments was that the test phase was divided into two subphases. In the first test phase, the participants were told to base their decisions on what seemed more familiar or better. Immediately after this test, they were asked what criteria they used for making their decisions. The aim was to assess their awareness at the level of understanding. In the second test phase, they were told that there was a rule that determined which of the alternatives was correct. They then repeated the test items from the first test phase, in a somewhat different order, but this time with a view to working out what this rule was. Again, no feedback was provided. The purpose of this second test phase was to see if participants who did well on generalization items in the first test phase were more likely to become aware of the rules when instructed to search for them in the second test phase. This would provide evidence that implicit learning can facilitate subsequent explicit learning. Mathews et al. (1989) found that participants who first performed an implicit training task were more likely to become aware of the rule system in a subsequent rule discovery task than those who did not, and they suggested that there might be a synergistic relationship between implicit and explicit learning processes. After this second test phase, the participants were again asked what criteria they used for making their decisions. If they did not mention animacy, the system was explained to them and they were asked whether they had become aware of the potential relevance of animacy at any point during the previous training or test tasks. Although the use of retrospective self-report is clearly not ideal (see Leow, 1997, for discussion), given the simplicity of the present system and the ease of conceptualizing the rules, it was assumed that participants would be able to recall whether any relevant hypotheses had occurred to them during the preceding tasks.

EXPERIMENT 1

Participants

There were 41 participants, with a mean age of 24.9 years. Sixty-eight percent of them were female. Eighty-eight percent were undergraduate or graduate students at the University of Cambridge. Thirty-four percent were nonnative speakers of English. Their L1s were Cantonese (*n* = 1), Dutch (*n* = 1), Greek (*n* = 4), Mandarin Chinese (*n* = 1), Portuguese (*n* = 1), Serbian (*n* = 4), Slovenian (*n* = 1), and Taiwanese (*n* = 1). Two participants had received equal exposure to two languages from birth: English and Malay in one case and Taiwanese and Mandarin in the other. Apart from English (which was advanced in all cases), the L2s known by the participants to an intermediate level or better were Cantonese (*n* = 1), French (*n* = 23), German (*n* = 15), Greek (*n* = 1), Ancient Greek (*n* = 1), Hokkien (*n* = 2), Italian (*n* = 5), Japanese (*n* = 1), Latin (*n* = 6), Malay (*n* = 1), Mandarin (*n* = 3), Russian (*n* = 1), Serbian (*n* = 1), and Spanish (*n* = 3). Thirty-four percent of the participants were studying language-related disciplines (linguistics, applied linguistics, modern languages). Participants from outside the author's department were paid £5.

Method

Materials. The noun phrases used for training and testing are shown in Table 1. The system is similar to that used in Williams (2004, Experiment 2), but in this version, there were twice as many nouns per class. The four novel words *gi*, *ro*, *ul*, and *ne* were selected so as to be unlike any determiners in languages that the participants were likely to know. They were assigned animacy values in such a way as to minimize confounds with sound. The words containing front vowels (*gi* and *ne*) expressed opposite values of animacy and distance, as did the two words containing back vowels (*ro* and *ul*). If, for instance, *gi* and *ne* had both been assigned to the animate category and *ro*

Table 1. Items used in Experiment 1

Living		Nonliving	
Near	Far	Near	Far
gi dog	ul dog	ro sofa	ne sofa
gi mouse	ul mouse	ro cup	ne cup
gi cow	ul cow	ro television	ne television
gi cat	ul cat	ro book	ne book
gi flies	ul flies	ro cushions	ne cushions
gi snakes	ul snakes	ro plates	ne plates
gi pigs	ul pigs	ro boxes	ne boxes
gi bears	ul bears	ro pictures	ne pictures
gi lion(s)	*(ul lion/s)*	ro table(s)	*(ne table/s)*
gi bird(s)	*(ul bird/s)*	ro vase(s)	*(ne vase/s)*
(gi monkey/s)	ul monkey(s)	*(ro stool/s)*	ne stool(s)
(gi bee/s)	ul bee(s)	*(ro clock/s)*	ne clock(s)

Note. Items in italics and parentheses were not presented during training but were withheld for testing generalization ability. Items used for the test of memory for trained items are underlined.

and *ul* to the inanimate category, then animacy would have been confounded with vowel type. As it was, any such biases worked in the opposite direction to animacy.

Nouns were selected so as to be clear cases of animate or inanimate objects. Natural things that do not typically move (e.g., trees or rocks) and man-made things that do move (e.g., bicycles or cars) were not used. What is referred to here as *animacy* could also be described as living versus nonliving, natural versus man-made, or moves versus does-not-move. There were 12 nouns of each class. Eight nouns of each class occurred with both possible determiners during training but in either singular or plural form (see Table 1). The remaining four nouns of each class only occurred with one determiner during training. These will be referred to as *critical* items, and the withheld combinations will be referred to as *generalization* test items. So that critical items occurred as often as the other items during training, they occurred in singular and plural forms (see Table 1). Across the training set, singular and plural nouns appeared equally often with each type of determiner.

Sentence contexts were written for each noun phrase from the training set. The sentences were intended to make it clear how far away the object was in the context of the situation described. A complete listing of the training sentences appears in Appendix A. The training items were divided into two sets, referred to as Set 1 and Set 2. Each noun appeared once in each set, but with either a different determiner or, for the critical items, with a different value of singular or plural. Each set contained all the nouns from each class, including all four critical items.

The materials for the test phase comprised the singular and plural forms of the withheld items from the training phase plus an additional 10 items (5 from each class) that had been presented during training (see Table 1). These will be referred to as generalization and trained test items, respectively. Sentence contexts were written for the test items; these sentences contained verbs that did not occur in the training items. The target noun phrase occurred at the end of the sentence.

After testing a number of participants, it was decided that it would be more prudent to ensure that the general situations described by the test and training sentences were as dissimilar as possible. For example, two of the original test items used the verbs *repair* and *fix*, and the correct response was *ro*. In the training set, *ro* had occurred once in the context of *restore* and once in the context of *broke*. It was decided to revise some of the test items so that the events they described were totally dissimilar to any events in the training set. These will be referred to as the *revised test items*. Some changes were also made to other items to strengthen the context. All of the test items are presented in Appendix B.

A second version of the training set was constructed, in which the assignment of the determiners to nouns was changed so that *gi* and *ul* occurred with inanimates and *ro* and *ne* occurred with animates. The same sentence contexts were used. This was to control for the possibility that certain sound combinations might be preferred when participants are forced to choose between alternatives in the test phase. For example, participants might prefer *ne lion* over *ul lion* simply because they do not like the combination of an *l*-final determiner and an *l*-initial noun. The test items were identical to those in the first version. Half of the participants were tested on each version.

Procedure.

Language background questionnaire. The participants first filled out a questionnaire concerning their age, field of study, L1, and knowledge of L2s. For L2s, the questionnaire elicited information concerning instructional context, qualifications received, contexts of use, and a self-assessment of proficiency (beginner, intermediate, upper intermediate, or advanced).

Phonological short-term memory test. The same PSTM test was used as in Williams and Lovatt (2003) and Williams (2004). This was an immediate serial recall task in which lists of three nonsense words were presented (both auditorily and visually) and the participant had to recall the nonsense words in each list in the order of presentation. The lists were composed from a pool of eight nonsense words: *johombe*, *zabide*, *wakime*, *migene*, *shosane*, *tisseke*, *chakume*, and *nawase*. There were 10 practice lists using different nonsense words (including the determiners used in Williams, 2004), followed by 13 experimental lists. One mark was awarded for each word recalled correctly and in the correct list position; half a mark was awarded for a single phoneme error or for a position error.

Vocabulary pretraining. Participants were presented with four cards that had a novel word printed on one side (e.g., *gi*) and an English word printed on the other (e.g., *near*). To distinguish the alternative forms for each meaning, the English words were printed in different colors (red, blue, gray, and green for *gi*, *ro*, *ul*, and *ne*, respectively). The participants studied the cards until they were able produce the novel word when given an English word in a certain color (e.g., to produce *gi* when presented with *near* in red). They then practiced this task in a computerized version until they could perform without error. Following this, the participants performed a comprehension task on the items. The computer played a recording of the author saying one of the words (e.g., *gi*), and the participants had to indicate whether it meant “near” or “far” by pressing one of two keys on the keyboard. A beep signaled an error. It was pointed out to them that the color scheme used in the previous task was now irrelevant, as it was only necessary to classify the words as having been paired with “near” or “far.” There were 12 repetitions of each item presented in a random order.

Training task. The participants were provided with written instructions that explained the general purpose of the experiment as follows:

> This is an experiment about how people formulate sentences in different languages. Different languages have different ways of saying things, which raises the question of whether this forces people to think slightly differently in order to speak and understand in different languages. Here I am

> investigating this issue in a situation where the sentences are almost entirely in English, apart from the four words that you have just learned.

The function of the novel words was then described in the context of the example sentence *The little boy patted gi tiger in the zoo* (note that this is not one of the training items):

> In this language, each time an object is mentioned, it is necessary to specify how far away it is from the subject of the sentence, whether it is "near" or "far." The word that is used to do this also functions like the English word "the." So, saying *gi tiger* is like saying "the-near tiger." In the context of this example, the word expresses the idea that the boy is near to the tiger (obviously, because he is patting it).

They were then told that their tasks were (a) to listen to each sentence, and on hearing one of the novel words, to indicate whether it means "near" or "far" by pressing the appropriate key; (b) after each sentence, to repeat it exactly aloud; and (c) during (a) and (b), to form a mental image of the general situation described by the sentence, paying particular attention to the distance relationship expressed by the novel word. They were told how many different sentences there would be, with three repetitions of each, and that there would be a memory test for some of the sentences at the end. However, it was stressed that it was not necessary to remember the exact wording of the sentences; it would be sufficient to understand the general situation described by the sentences, which was why it was important to form a mental image as they listened to them. These instructions were designed to deter the participants from trying to remember the precise forms of the novel words and to concentrate instead on their meanings and how they contributed to their understanding of the overall situation described by the sentence. In the course of running the experiment, it was found that some of the nonnatives found it arduous to repeat the sentences in their entirety. In such cases, they were instructed to repeat only the relevant noun phrase (e.g., *gi lion*) while still forming an image of the overall situation.

The order of trials within each set was independently randomized for each participant. The training phase consisted of six blocks of 24 trials in the order Set 1, Set 2, Set 1, Set 2, Set 1, Set 2. There was a rest break for a minimum of 30 seconds between each block. Because of the distribution of items over sets, the same noun never occurred in consecutive trials, thereby reducing the possibility of making explicit comparisons between items.

Testing phase. After the training phase, the participants were told that before performing the memory test, they would have to perform an exercise on new sentences. For each test item, the computer first displayed the sentence context (e.g., *The lady spent many hours sewing . . .*). Upon clicking a *continue* button, two alternative completions were presented, one above the other (*gi cushions / ro cushions*). The participants were told to select the noun phrase that seemed "more familiar, better, or more appropriate" on the basis of what they had heard during the training task. The correct phrase occurred as the upper or lower choice with equal frequency.

The testing items were presented in the same order for all participants. This order was contrived in order to reduce the possibility of making comparisons between successive items (e.g., a "far" animate followed by a "far" inanimate). The first test began with two trained items, followed by eight generalization items. Each generalization noun occurred once, and there were equal numbers of singular and plural nouns with each determiner. There were then eight more trained items, followed again by the eight generalization items, but this time, the generalization items showed the opposite value of singular or plural and were in different sentence contexts. Appendix B lists the test items in the order of presentation in this test.

After completing the first test phase, the participants were asked what criteria they had used to make their choices. Any references to living or nonliving, moves or does-not-move, and so forth were interpreted as evidence of awareness of animacy. In such cases, the participants were asked at what point in the experiment they had become aware that animacy was relevant. Participants who did not mention animacy were then told that in fact there was a rule that determined the correct choice and that they were to go through the test items again, but this time, with a view to working out what the rule was. It was explained that this time, the test would begin with items that should seem more familiar and that they should therefore be able to use their responses to these items as a basis for forming hypotheses. This second test phase began with 10 trained items, followed by the first set of generalization items, and then the second set of generalization items. The same sentence contexts were used as in the first test phase. Afterward, the participants were asked for their ideas about what the rule might be; if animacy was mentioned, they were then asked at what point in the experiment they had become aware of its relevance. If they did not mention animacy, the system was explained to them, and they were asked if at any point during the experiment they had considered the possible relevance of animacy.

Memory test. A yes/no sentence recognition test was used as the memory test. There were 10 sentences from the training set and 10 new sentences. The items were presented by computer and in a different random order for each participant.

The entire procedure took about 1 hour.

Results

A total of 8 of the 41 participants expressed awareness of the relevance of animacy when interviewed at the end of the first test phase. They were also able to state the specific relationships between the determiners and noun animacy. Six of them had become aware of the rule during the training phase, and two of them had only worked out the rule during the first testing phase.[2]

As one would expect, their generalization test performance was almost perfect (91.4% averaged over the two sets of generalization items in the first test phase, and significantly above chance, $p < .001$).

Of the remaining 33 participants, 6 described some kind of rule, but with low levels of confidence. None of them mentioned animacy, but rather referred to the syntactic structure of the sentence, the phonology of the noun, the nature of the action (without being more specific), or the directness of contact involved in the action (and, in one case, also the visibility of the object in the case of "far" items). The remaining 27 claimed to have based their decisions on familiarity or intuition. When told what the rule was, all 33 participants claimed not to have considered the relevance of animacy at any point during the experiment. They were, therefore, classified as unaware with respect to the target rule (i.e., as having no awareness at the level of understanding). Table 2 shows their mean scores in the first test phase, grouped by the components of the test in the order that they occurred in the experiment (with the exception of the first two trained items; these have been combined with trained 1). The difference between scores in each condition and chance (50%) was evaluated using single group *t*-tests. The mean over the two sets of generalization items in the first test phase is also shown (mean generalization). All test scores were significantly above chance, even for generalization items, where neither alternative had, in fact, been presented during training.[3]

Interestingly, the six participants who reported non-animacy-based rules performed poorly, with a mean generalization test score of 49%. In contrast, they scored 68% on trained items, which is significantly above chance, $t(5) = 3.84$, $p < .05$, and similar to the level of performance of the overall sample. Thus, the generalization test scores in Table 2 are actually being depressed by the inclusion of the participants who reported erroneous rules. If they are excluded, the generalization scores become: generalization 1, 65%, $p < .001$; generalization 2, 62%, $p < .01$; mean generalization, 63%, $p < .001$.

Recall that the original set of generalization test items contained a few verbs that were semantically similar to verbs that had occurred during training, whereas in the revised test, all of the generalization items used verbs that

Table 2. Mean percentage correct, standard deviations, and *t*-values for deviation from chance (*t*): 33 unaware participants (Experiment 1, first test phase)

	Generalization 1	Trained 1	Generalization 2	Mean generalization
M	61.7	70.4	59.8	60.8
SD	21.2	13.7	19.4	16.0
$t(32)$	3.18*	8.51**	2.91*	3.87**

*$p < .01$
**$p < .001$

were dissimilar to those used in training. There was no evidence that this change in the materials influenced the results. For the 18 of the 33 unaware participants who received the original test, the mean generalization test score was 62.1%, which was significantly above chance, $p < .01$. For the 15 participants who received the revised test, the mean generalization score was 59.2%, which was again significantly above chance, $p < .05$.[4] Also, there was no effect of the language version that the participants received. For example, for the 16 participants who received the version in which *gi* and *ul* were animate, the mean generalization performance was 59% (significantly above chance, $p < .05$), and for the 17 participants who received the version in which *ro* and *ne* were animate, the mean generalization performance was 62.5% (significantly above chance, $p < .01$).

It is possible that the above-chance level of responding to generalization items was just due to a few items where the sentence context perhaps reminded the participants of a relevant training sentence. On the other hand, if the effect represents an animacy bias, then it should be present for all of the test items. Therefore, test performance was analyzed by items and was found to be significantly above chance for the 16 generalization items, $p < 0.001$, and for the 10 trained items, $p < .001$. Of the 16 generalization items, 13 had mean scores above chance, 2 were at 50%, and 1 was less than 50%. Clearly then, the effects were not due to just a few items.

In the second test phase, the unaware participants were instructed to search for a rule that determined the use of the novel words. Following this exercise, a total of 11 participants were aware of the system, and the remaining 22 were either unable to report a rule or suggested other rules with low levels of confidence. The scores for these two groups over both test phases are compared in Table 3.

As would be expected, the participants who were aware of the system at the end of this second test phase had significantly higher scores for the trained (trained 2) and generalization (generalization 3) items than those who were not aware, as confirmed by independent sample *t*-tests (see the bottom row of Table 3). However, the performance for the 22 participants who were still unaware of the rule at the end of this phase was still significantly above chance for both trained items and generalization items.

It is now interesting to look back at the generalization test performance for these two groups in the first test phase. Did the participants who became aware of the rule in the second test phase show stronger evidence for implicit learning in the first test phase? There was some evidence of this for the first set of generalization items, but the difference between the aware and unaware groups does not quite reach significance, $p = .058$. Single group *t*-tests showed that at generalization 1, only the score for the aware group was significantly above chance. There is some evidence here, therefore, that those participants who became aware in the second test phase showed more evidence for implicit learning in the first test phase—at least when assessed by the first set of generalization items. Note, however, that even participants who remained

Table 3. Performance of the participants who were unaware of the rule after the first test phase of Experiment 1 according to whether they expressed awareness after the second test phase

Group	First test phase				Second test phase	
	Generalization 1	Trained 1	Generalization 2	Mean generalization	Trained 2	Generalization 3
Aware after second test phase (n = 11)						
M	71.6	64.5	59.1	65.3	80.9	88.6
SD	18.6	16.3	14.9	12.3	15.1	17.0
t(10)	3.85**	2.95*	2.02^{+}	4.14**	6.77***	7.56***
Unaware after second test phase (n = 22)						
M	56.8	73.3	60.2	58.5	63.8	59.2
SD	21.0	11.6	21.7	17.4	15.0	15.4
t(21)	1.52	9.43***	2.21*	2.29*	4.22***	2.75*
t(31) of difference	1.97^{+}	1.78^{+}	0.156	1.16	3.05**	4.96***

Note. The table shows mean percentage correct, standard deviations, and t-values for deviation from chance (t). The significance of each between-group difference (t of difference) is also shown.
$^{+}p < .1$
$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$

unaware at the end of the second test phase showed above-chance generalization test performance in the first test phase, at least for the second set of items, and when the mean of the two sets of generalization items is considered..

For the overall sample of participants, the mean score on the PSTM test was 70%, and the mean number of gender languages known to an intermediate level or better was 1.68 ($SD = 1.29$). The analysis of individual differences first focused on the mean generalization performance in the first test phase for the 33 participants who claimed to be unaware at that point. The performance of the 11 participants who spoke a gender L1 was not significantly different from the 22 participants who did not [64% vs. 59%, $t(31) = 0.86$]. However, performance for the 12 participants studying language-related disciplines—the *linguists*—was significantly better than for the 21 nonlinguists [69.3% vs. 55.9%, $t(31) = 2.47$, $p < .05$]. Next, correlations between mean phase 1 generalization performance and the following factors were calculated: performance on the PSTM test, $r = .01$, the number of L2s spoken to an intermediate level or better, $r = .263$, and the number of gender languages spoken to an intermediate level or better, including L1, $r = 0.357$, $p < .05$. Therefore, both studying linguistics and knowing gender languages—but not necessarily as an L1—were associated with better performance. However, these two factors are also somewhat related given that the linguists knew more gender languages than the nonlinguists (a mean of 2.17 as opposed to 1.48), although the difference was not significant, $t(31) = 1.55$.

Discussion

At the end of the first test phase, the majority (80%) of the participants claimed to have had no awareness of the rules, and not even of the relevance of animacy, during either the training or test phases. However, these unaware participants were able to select the correct determiner-noun combination at significantly above-chance levels, even for generalization items where neither alternative had been presented during training and the contexts in which the words occurred were novel. The majority of these participants (66%) also failed to become aware of the relevance of animacy when they were instructed to search for a rule in the second test phase. Yet, their performance on generalization test items was still significantly above chance and was at about the same level as it had been in the first phase, prior to the instruction to search for rules. There seems good reason to believe, therefore, that these participants really had no awareness of the relevance of animacy. However, their responses showed a small but statistically significant animacy bias. It appears, then, that form-meaning connections were learned under conditions where only the relevant forms were noticed in the input. The participants were not aware of the relevant aspect of meaning and did not integrate it with the relevant forms at the time of encoding. Furthermore, whatever processes were responsible for distilling the animacy rule from the training sentences must have been operating implicitly, outside of the participants' awareness.

With regard to possible synergies between explicit and implicit learning, it was found that participants who became aware after the instruction to search for rules tended to show stronger implicit learning in the first test phase, at least as assessed by the first set of generalization items. Because these occurred at the beginning of the test phase, it is not unreasonable to suppose that they were more likely to reflect intuitive responding, as opposed to later in the test, where hypothesis-formation activity might have begun to interfere with performance (recall that this group of participants was classed as unaware after the first test phase, and so any hypotheses that they formed would have been erroneous). Thus, there is some evidence in support of the notion that implicit learning can feed into explicit learning processes (Mathews et al., 1989).

However, there is an alternative explanation of the results. It could be the case that the participants who became aware after the second test phase in fact had some vague or partial level of awareness at the level of understanding during the first test phase, but they were unable to make their knowledge fully explicit at that point. If this were true, it would invalidate the data from this group. The only evidence against this argument is that the participants who were aware at the end of the second test phase all said that they were not aware of the relevance—or even potential relevance—of animacy prior to the instruction to search for rules. It is not unreasonable to suppose that, given the benefit of hindsight, these participants would have been able to report if any of their earlier vague or partial intuitions corresponded to their ultimate understanding of the system. However, even if one were to be sceptical about these participants' lack of awareness in the first test phase, it is still important to note that first-phase generalization performance was significantly above chance even for the participants who were still unaware after the second test phase.

With regard to individual differences, it was found that for the participants who were unaware after the first test phase, better generalization test performance was associated with knowing gender languages and studying a language-related subject. It is not clear at the moment which of these factors is the more relevant (see Experiment 2 for further evidence), but it does appear that knowledge factors had some effect. Phonological short-term memory was not related to learning in this experiment. I will return to the issue of memory factors in the General Discussion section.

Having obtained evidence for implicit learning in Experiment 1, we must now consider more closely the nature of what was learned. The aware participants described the system in terms of form-meaning relationships, as a mapping between determiners and values of noun animacy. However, it is logically possible that implicit learning was in fact driven by a statistical analysis of the distributional structure of the input, an analysis that made no reference to meaning at all. For example, participants could have simply learned that any noun that takes *gi* also takes *ul*, and any noun that takes *ro* also takes *ne*. This is because the majority of the nouns in each category appeared with both possible determiners during training.

There are two ways that the contribution of distributional analysis could be assessed. One way would be to construct a version of the language in which the distributional structure is preserved but the correlation between noun class and animacy is removed. However, previous research has failed to find evidence of learning noun class systems in the absence of phonological or semantic cues, either under intentional (Braine et al., 1990; Brooks, Braine, Catalano, & Brody, 1993; Frigo & McDonald, 1998) or implicit (Williams, 2003) learning conditions, so the likely outcome of such an experiment would be chance level responding to generalization test items. The alternative is to remove the distributional structure by ensuring that each noun only occurs with one possible determiner during training. This second option was preferred because it would confirm implicit learning of form-meaning connections through a positive, rather than a null, result.

EXPERIMENT 2

Participants

There were 24 participants, with a mean age of 27.4 years. Ninety-two percent of them were female. They were all undergraduate or graduate students at the University of Cambridge. They were all advanced nonnative speakers of English.[5] Their L1s were Albanian ($n = 2$), Arabic ($n = 1$), Bosnian ($n = 1$), Mandarin Chinese ($n = 3$), Croatian ($n = 2$), German ($n = 2$), Korean ($n = 1$), Macedonian ($n = 2$), Serbian ($n = 1$), Singhalese ($n = 3$), Spanish ($n = 1$), Taiwanese ($n = 1$), and Ukrainian ($n = 1$). There were three bilinguals: Russian-German, Tamil-English, and Cantonese-English. L2s known to an intermediate level or better were French ($n = 6$), German ($n = 1$), Italian ($n = 5$), Latin ($n = 3$), Mandarin Chinese ($n = 2$), Polish ($n = 1$), Russian ($n = 1$), and Serbian ($n = 2$). Fifty-four percent were studying language-related disciplines (e.g., linguistics, applied linguistics, or modern languages). Participants from outside the author's department were paid £5.

Method

Materials and Design. The noun phrases used in the experiment are shown in Table 4. The critical items, generalization items, and trained test items were all the same as in Experiment 1 so that exactly the same test material could be used. The only difference was that during training, each noun only occurred with one determiner. So that there would be the same number of different sentences as in Experiment 1, each noun occurred in both singular and plural forms and in a different sentence context in each case. The materials were

Table 4. Items used in Experiment 2

Living		Nonliving	
Near	Far	Near	Far
gi dog(s)	ul mouse(s)	ro cushion(s)	ne book(s)
gi cow (s)	ul cat(s)	ro sofa(s)	ne television(s)
gi fly(s)	ul pig(s)	ro cup(s)	ne picture(s)
gi snake(s)	ul bear(s)	ro box(s)	ne plate(s)
gi lion(s)	*(ul lion/s)*	ro table(s)	*(ne table/s)*
gi bird(s)	*(ul bird/s)*	ro vase(s)	*(ne vase/s)*
(gi monkey/s)	ul monkey(s)	*(ro stool/s)*	ne stool(s)
(gi bee/s)	ul bee(s)	*(ro clock/s)*	ne clock(s)

Note. Items in italics and parentheses were not presented during training but were withheld for testing generalization ability. Items used for the test of memory for trained items are underlined.

divided into two sets, as for Experiment 1, so that each noun occurred once in each set and with half singular and plural forms in each set. The items and sentence contexts in the first set were the same as those used in Experiment 1. New sentence contexts were written for the items in the second set (see Appendix A).

A second version of the training set was constructed in which the assignment of the determiners to nouns was changed so that *gi* and *ul* occurred with inanimates and *ro* and *ne* occurred with animates. Equal numbers of participants received each version. The testing material was exactly the same as the revised version of Experiment 1. The test sentences contained different verbs and described different situations from the training sentences.

Procedure. The procedure was identical to Experiment 1. The experiment again lasted about 1 hour.

Results

Seven of the 24 participants expressed awareness of the relevance of animacy and were able to describe the rules when interviewed at the end of the first test phase. Two of them had become aware during the training phase, and the remaining five reported that the rule "just came to them" during the first test phase. The mean first test phase generalization performance for this aware group was 85%, significantly above chance, $p < .01$.[6] It is interesting that for the five participants who became aware during the test phase, performance over the first set of generalization items, constituting the first third of the test, was 75%, significantly above chance, $p < .05$. Either rule awareness devel-

Table 5. Mean percentage correct, standard deviations, and *t*-values for deviation from chance (*t*): 17 unaware participants (Experiment 2: first test phase)

	Generalization 1	Trained 1	Generalization 2	Mean generalization
M	64.0	84.7	64.0	64.0
SD	17.6	11.1	16.5	14.2
t(32)	3.27*	12.88**	3.50*	4.04**

*$p < .01$
**$p < .001$

oped very quickly or else accurate responding by intuition led to subsequent rule awareness.

The scores for the remaining 17 unaware participants in the first test phase are shown in Table 5. Significant differences from the chance level of performance are also indicated. All scores were well above chance, even for the generalization items.[7] Items analyses were also performed. The mean score over the 16 generalization test items was significantly above chance, $p < .001$, as was the score over the trained items, $p < .001$. Of the 16 generalization items, 13 had above-chance scores, 1 was slightly below chance, and 2 were at chance. Only one of these items was also at chance in Experiment 1.

In the second test phase, only three of the previously unaware participants were able to correctly state the rule. This made it impossible to make meaningful comparisons in performance over the first test phase. The situation was also complicated by the fact that a further three participants were classed as *partially aware* because they had realized the relevance of animacy but only stated the correct relationship to animacy for one pair of determiners (i.e., either the “near” or “far” pair). Their performance was poor over all three sets of generalization test items (54%, 46%, and 48%). What was clear, however, was that phase 1 generalization test performance was well above chance for the remaining 11 participants who still claimed to be unaware after the end of the second test phase. The data are shown in Table 6.

For the whole sample, the mean score on the PSTM test was 71% and the mean number of gender languages known to an intermediate level or better was 1.42 ($SD = 1.47$).

The analysis of individual differences focused on the mean phase 1 generalization test performance for the 17 participants who claimed to be unaware at that point. This time, the performance of the eight participants who spoke a gender L1 was significantly better than the nine participants who did not, 72% versus 57%, $t(15) = 2.48$, $p < 0.05$, although participants who spoke a gender L1 also knew more gender languages overall than those who did not—a mean of 2.0 versus 0.11, $t(15) = 5.05$, $p < .001$. This time, performance for

Table 6. Mean percentage correct, standard deviations, and *t*-values for deviation from chance (*t*): 11 unaware participants (after Experiment 2, second test phase)

	First test phase			Second test phase		
	Generalization 1	Trained 1	Generalization 2	Mean generalization	Trained 2	Generalization 3
M	65.9	83.6	68.2	67.0	70.0	59.4
SD	13.8	9.0	16.2	14.2	25.3	12.6
$t(10)$	3.82**	12.44***	3.73**	4.89***	2.62*	2.49*

*$p < .05$
**$p < .01$
***$p < .001$

the 10 linguists was not significantly better than for the 7 nonlinguists, 66% versus 61%, $t(15) = 0.78$. Correlations between mean phase 1 generalization performance and the following factors were calculated: PSTM, $r = .228$, the number of L2s spoken to an intermediate level or better, $r = .404$, and the number of gender languages spoken to an intermediate level or better, including L1, $r = .358$. However, it must be recognized that these correlational analyses are highly unreliable because of the small sample size and the limited variation in the number of gender languages known by the unaware participants ($M = 1.0$, $SD = 1.22$). Nevertheless, the fact that participants who spoke a gender L1 outperformed those who did not does suggest that—just as in Experiment 1—knowledge of gender languages had some effect.

Discussion

As in Experiment 1, there was strong evidence of an animacy bias in responding, despite a lack of awareness of the relevance of animacy. Seventeen of the participants did not express any awareness of the relevance of animacy at the end of the first test phase, yet they were significantly better than chance at choosing the correct determiner for generalization items. When these participants were instructed to search for a rule in the second test phase, 11 of them were unable to discover the relevance of animacy, but their responses to generalization items were still above chance. Furthermore, looking back at the first test phase, the generalization responses for these 11 participants were also significantly above chance.

Test scores were—if anything—slightly better than those of the unaware group in Experiment 1. In the case of trained items, this presumably reflects the fact that, ignoring singular and plural variants, each determiner-noun combination occurred twice as often in this experiment. Whether this improved memory for trained items also influenced generalization performance is not clear. What is important is that generalization performance was at least as good as that in Experiment 1 despite the fact that each noun only ever occurred with one determiner during training. Participants were able to learn a direct association between determiners and animacy without the support of distributional information.

With regard to individual differences, as in Experiment 1, knowledge of gender languages was associated with better phase 1 generalization test performance. In this case, however, whether or not the L1 was a gender language was the main factor. Speakers of gender L1s also differed markedly in terms of the number of gender languages that they knew, but the correlation between number of gender languages and generalization score was not significant.

GENERAL DISCUSSION

The present experiments show that, at least for some individuals, it is possible to learn form-meaning connections without awareness of what those connections are. This final section will discuss the issue of levels of awareness, the relationship between these and earlier results, individual differences, and, finally, whether the present results can be generalized to other types of form-meaning connection.

Levels of Awareness

In this study, learner awareness was assessed through oral interviews conducted by the experimenter after the first and second test phases. The purpose of these interviews was to establish whether the participants were aware that noun animacy was relevant to the use of the determiners. Clearly, however, the use of posttask interviews is less than satisfactory. Is it not possible that participants had some kind of fleeting awareness of the relevance of animacy during the tasks but did not report it? In fact, it has been suggested that the contents of awareness might actually extend beyond that which is attended to and which is reportable and that there might be a kind of fleeting *phenomenal awareness*, the contents of which are rapidly forgotten (Block, 2001; Lamme, 2003). This would be distinct from *access* or *reflective* awareness—corresponding to focal attention—the contents of which are reportable. In this sense, participants might have had phenomenal awareness of noun animacy or even its relevance to the novel determiners during the present tasks. We have no way of knowing. What we do know is that unaware participants did not knowingly use these fleeting impressions to guide their responses in the test tasks, nor did they have any recognition for the rule when they did become aware of it. When they were told what the rule was or they had discovered the rule when invited to search for it in the second test phase, they claimed not to have been aware of the relevance of animacy prior to that point. Whatever phenomenal awareness they might have had appeared to have been immediately forgotten and did not form the basis for rule learning or intentional behavior. Therefore, even with the possibility of phenomenal awareness in mind, it is still possible to claim that learning was implicit.

Leow (1997) identified another level of awareness, a subdivision of awareness at the level of noticing into ±*meta-awareness*. In his data, noticing without meta-awareness occurred when participants simply corrected a verb stem from, say, *repetió* to *repitió* without comment. Meta-awareness seemed to involve noticing the contrast between one form and another—for example, commenting on the fact that *repetir* and *repitió* contained different stem vowels. Such comments were limited to specific examples, and there was no statement of a rule (i.e., that *e* changes to *i*). In the present tasks, participants classified as unaware might have noticed that different forms were being used for the same function (e.g., the use of *gi* in one sentence and *ro* in another) without identifying the reason why. This is not unlikely because the near-far judgment task involved remembering that pairs of words were associated with different responses, and the test tasks involved choosing between two forms for the same meaning. However, remarkably few participants reported having spontaneously looked for a rule that would explain the alternation of forms during the training task (five in Experiment 1, two of whom were classified as unaware, and two in Experiment 2, both of whom were classified as aware). However, what is not known is whether greater awareness of the alternating forms during training was associated with greater implicit learning for the unaware group. For example, participants might say to themselves that "*gi* was used in that sentence instead of *ro*." This would constitute a kind of meta-awareness of determiner usage that might well have facilitated learning—if for no other reason than it improved item memory for the trained determiner-noun combinations. However, it would not have been picked up in the posttask interviews. Clearly, a finer-grained approach to assessing learner awareness would have been preferable, ideally utilizing think-aloud procedures during the training and test tasks.

Relationship to Previous Results

The present results differ from those obtained in Experiment 2 of Williams (2004) where no implicit learning was obtained, even though a similar system was used in both cases. The main difference between the experiments is that the present training task was less demanding for the participants because it required comprehension as opposed to sentence composition. Thus, there might be a negative relationship between task demands and implicit learning. As noted previously, even though there are numerous demonstrations of reduced implicit learning under dual-task conditions (Carr & Curran, 1994; Frensch, Wenke, & Runger, 1999), it is still not clear precisely why this occurs.

On the basis of their results, Jiménez and Méndez (1999) argued that one cannot simply appeal to a reduction in processing resources. One alternative possibility that they suggested is that secondary tasks interfere with implicit learning because they upset the temporal relationship between the critical stimuli. It has been found that if associations between stimuli are to be learned, then they have to be in close temporal contiguity (Goshen-Gottstein & Moscovitch, 1995). In the case of form-meaning connections, this could be achieved by intentional integration of the stimuli at encoding, as stressed by Doughty and Williams (1998). However, if associations are to be learned without awareness, it is presumably important that the stimuli just happen to be processed more or less simultaneously. The more complex the task environment and the more other things the learner has to think about, the less likely this will be.

However, the present experiments also differ from the earlier one in terms of the range of exemplars provided in training, as there were twice as many nouns in the present case. The difference in results could therefore also be a reflection of a critical mass effect (Marchman & Bates, 1994; Robinson & Mervis, 1998). Clearly, both of these possibilities are worthy of further investigation, but they lie beyond the scope of the present study.

Individual Differences

Knowledge of gender languages appeared to be related to implicit learning in these experiments. In Experiment 1, implicit learning correlated—albeit fairly weakly—with the number of gender languages known. In Experiment 2, participants who spoke a gender L1 did better than those who did not, although, unsurprisingly, participants who spoke a gender L1 also knew significantly more gender languages. Implicit learning was not significantly correlated with the number of gender languages known, but the limited sample size must be borne in mind here. In Williams (2004, Experiment 1), implicit learning was correlated with the number of gender languages known, and there was also a gender L1 effect. It is important to note that in all these studies, it was specifically the knowledge of gender languages that was relevant, given that in no case was there a significant correlation with the overall number of languages known by the participants. Thus, despite efforts in the present experiments to make the target system appear less like the kind of noun class and determiner systems with which the participants were familiar, an influence of gender language knowledge was obtained, although it remains unclear to what extent having acquired that knowledge in the L1 is important. I consider why gender language knowledge could have aided learning in the following subsection.

The other individual differences factor that was considered here was PSTM, but in neither of the experiments was it even mildly related to learning outcomes. This contrasts with the results of Williams and Lovatt (2003), where PSTM was related to learning a noun class system independently of the effect of gender language knowledge. However, the Williams and Lovatt system involved an arbitrary noun class distinction, signaled only by the distribution of the determiners. Learning depended on remembering precisely which determiners had occurred with which nouns (i.e., on remembering form-form associations). This would potentially place great demands on phonological memory. Similarly, in Williams (1999), correlations with memory measures were obtained for inductive learning of surface agreement—specifically, euphony among modifiers, articles, and nouns in Italian phrases like *la musica moderna*. In contrast, there were no correlations between memory measures and learning the functions of noun inflections (singular and plural) and verb inflections (person and number). It would appear, therefore, that although learning about the distribution of forms is dependent on phonological memory, learning about the correlations between forms and meanings is not.

Generalizing Beyond Word Classes

The present results appear to run counter to the common belief that learning form-meaning mappings requires attention to both form and some hypothesized meaning at encoding and that discovering the conditions on the use of a word requires explicit hypothesis-testing processes. More generally, the results also seem to contradict the assumption that awareness at the level of noticing is a necessary condition for learning (Schmidt, 1990, 1994, 2001). Instead, they appear to support Tomlin and Villa's (1994) contention that stimuli that are processed outside of awareness—in their terms, *detected*—can contribute to learning. Simard and Wong (2001) criticized the empirical basis of Tomlin and Villa's argument for the sufficiency of detection because it was based either on evidence of momentary activation of familiar stimuli (as in subliminal priming experiments) or artificial grammars and event sequences that are of dubious relevance to natural language. They suggested that in order for Tomlin and Villa's argument to be relevant to SLA, it would have to be shown that, for example, form-meaning connections could be made below the threshold of awareness and influence subsequent task performance. It is tempting to see the present results as providing the kind of proof that Simard and Wong called for: The animacy feature was detected, or activated, outside of awareness when the input sentences were processed, became associated with the determiner in the input sentence, and influenced subsequent generalization test performance.

However, it would be premature to draw general conclusions about the role of awareness in language learning, or even learning form-meaning connections, from the present results. Claims for explicit learning of form-meaning mappings are generally made in relation to open-class words (nouns, verbs, adjectives, etc.), but the present experiments targeted determiners. It is possible that at an unconscious level, the animacy feature was not simply interpreted as a condition on the use of the determiners but as an essentially grammatical feature that defined grammatical noun classes and that was used to control agreement within the noun phrase. Such a grammatical construal of the problem, and the grammaticization of animacy, would make the present learning problem potentially very different from the acquisition of form-meaning mappings in open-class vocabulary learning.

The persistent relationship between generalization performance and knowledge of gender languages is consistent with the idea that grammatical knowledge was a relevant factor. If the learning process were the same as that used in general vocabulary learning, then why would such a relationship obtain? None of the participants spoke languages in which word classes were distinguished by animacy, so the effect cannot be explained in terms of a heightened sensitivity to this feature. The only thing that these languages share with the target system is a formal structure in which nouns belong to different classes and determiner use is controlled in part by agreement processes with respect to those classes. The fact that prior knowledge of such systems appeared to facilitate implicit learning does suggest that the input was being interpreted in terms of noun classes and agreement processes. Therefore, the present results could be taken as highlighting the importance of prior knowledge in helping learners to implicitly assimilate linguistic rules, and, so, they might not generalize to learning open class vocabulary.[8] Nevertheless, they might still be taken to demonstrate that information that is detected outside of awareness can enter into learning processes (Tomlin & Villa, 1994), at least in the domain examined here.

However, there is another consideration that threatens even this restricted role for learning through detection without awareness. In the present experiments, the relevant meaning feature was implicit in the meaning of a word in the sentence (the noun accompanying the determiner) rather than being derived from the sentence context. To see the importance of this distinction, suppose, for argument's sake, that participants had been introduced to *gi* and *ul* as variants of the definite article and to *ro* and *ne* as variants of the indefinite article and that the experiment examined whether they could implicitly induce a correlation with distance, with *gi* and *ro* being used for near objects and *ul* and *ne* for far. In this case, the relevant meaning feature would be present in the situation model (Kintsch, 1988) that the participant constructs in response to the entire sentence. Assuming that the sentences and the task were devised in such a way that the distance feature would be reliably represented in the situation model, if only implicitly, would its connection to the determiners be learned implicitly?

One possibility is that form-meaning connections are established online as sentences are processed. Whatever features are activated by the sentence are available for association with forms in the input, whether those features derived from specific words (as in the case of animacy) or from the situation model (as in the case of distance). In this view, there is no reason to suppose that the two kinds of feature would be differentially available. However, suppose that the form-meaning connection emerges through offline abstraction processes in memory. To learn a correlation with a distance feature, the determiner would have to be encoded in memory along with the situation model of the entire event, and the relevant feature would have to emerge as an associate of the determiner through comparisons across many events. To learn a correlation with an animacy feature, it is only necessary to encode the association between the determiner and the noun. Being a more circumscribed search space, one would expect the relevant feature to emerge more readily than if entire situations had to be compared. However, additionally, it might not even matter what meaning features of the noun had been activated and incorporated into the situation model when the original sentence had been interpreted. Assuming that offline abstraction processes operate over both the forms and the lexically stored meanings of the words, then it would be possible to derive the correlation between determiner use and animacy regardless of the meaning features that were activated in the context of the original sentence. Therefore, meaning features that are implicit in lexical representations of words should be more likely to contribute to implicit learning than features that derive from sentence interpretations because they do not have to be stored in memory when the input is processed but can be activated in memory later. The same argument would apply to other lexically represented properties of words that might correlate with noun class, such as phonology. Indeed, it is well documented that adult L2 learners show (over-) sensitivity to phonological cues to grammatical gender (Holmes & Dejean de la Batie, 1999; Taraban & Kempe, 1999).

Thus, there are a number of reasons why one needs to be cautious in generalizing from the present results. It is clearly an important empirical question whether implicit learning of form-meaning connections can be obtained in cases where the relevant feature is implicit in the situation model, rather than the meaning of words. However, what the present study suggests is that a meaning feature can contribute to implicit learning when it derives from the lexical meaning of a word, when it enters into a grammatical agreement process, and when the learner can draw on relevant grammatical knowledge. Implicit learning of form-meaning connections is possible, at least in principle.

NOTES

1. The reason for the discrepancies with patients such as H. M. appears to be related to the localization of the brain damage. Vargha-Khadem et al. (1997) proposed that patients with damage specific to the hippocampus (e.g., their infant amnesics) will show impaired episodic memory but normal semantic and vocabulary learning, whereas if there is also damage to surrounding cortical areas (as appears to have been the case for H. M.), then semantic learning will be impaired as well. Verfaellie et al. (2000) supported this argument by comparing their patient, PS, with another patient, SS, who had more diffuse damage in the hippocampal region and who showed impaired vocabulary learning as well as impaired episodic memory.

2. It was noted that of the six participants who became aware during the training phase, five were native speakers of English. This possibly reflects the lower processing demands of performing the training task in the native language, making it more likely that the rule will be noticed.

3. Despite showing above-chance performance, the unaware group performed significantly below the level of the aware group on all tests in the first test phase: generalization 1 (61.7% vs. 84.4%, $p < .05$), trained items (70.4% vs. 90.0%, $p < .01$), generalization 2 (59.8% vs. 98.5%, $p < .001$).

4. The participants who received the revised version knew slightly fewer gender languages than those in the original version (1.53 and 1.89, respectively). Given the positive correlation between knowledge of gender languages and learning (see discussion), this might explain the slightly lower level of learning for participants who received the revised version.

5. Nonnatives were used because there was some evidence from Experiment 1 that natives were more likely to become aware of the rule during the training task (see Note 2).

6. One participant in this group expressed awareness of the relevance of animacy but reported the wrong mapping onto the words for "far." Her mean performance on the generalization items was 44%. The mean over the remaining six aware participants was 92%.

7. Scores for the unaware group were significantly worse than the aware group, although only for the generalization items: generalization 1 (64.0% vs. 82.1%, $p < .01$), trained items (84.7% vs. 91.4%, $p = .20$), generalization 2 (64.0% vs. 87.5%, $p < .01$).

8. Hirst, Phelps, Johnson, and Volpe (1988) made a similar suggestion in relation to language learning in an amnesic woman with prior linguistic expertise.

REFERENCES

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition*, *79*, 197–219.

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

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

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences. *Studies in Second Language Acquisition*, *16*, 205–230.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37.

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., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.

Ellis, N. C. (1994). Vocabulary acquisition: The explicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). San Diego, CA: Academic Press.

Frensch, P. A., Wenke, D., & Runger, D. (1999). A secondary tone-counting task suppresses expression of knowledge in the serial reaction task. *Journal of Experimental Psychology-Learning Memory and Cognition*, *25*, 260–274.

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

Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following medial temporal lobe resection. *Brain & Cognition*, *7*, 157–177.

Goshen-Gottstein, Y., & Moscovitch, M. (1995). Repetition priming effects for newly formed associations are perceptually based: Evidence from shallow encoding and format specificity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1249–1262.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 447–481). Mahwah, NJ: Erlbaum.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64.

Hirst, W., Phelps, E. A., Johnson, M. K., & Volpe, B. T. (1988). Amnesia and second language learning. *Brain and Cognition*, *8*, 105–116.

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

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 236–259.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.
Kitchener, E. G., Hodges, J. R., & McCarthy, R. (1998). Acquisition of post-morbid vocabulary and semantic facts in the absence of episodic memory. *Brain*, *121*, 1313–1327.
Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 169–181.
Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*, 12–18.
Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*, 467–505.
Leow, R. P. (2000). A study of the role of awareness in foreign language behavior. *Studies in Second Language Acquisition*, *22*, 557–584.
Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1022–1050.
Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*, 339–366.
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.
McClelland, J., & Rumelhart, D. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.
Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition*, *79*, 115–134.
Merriman, W. E. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp. 331–358). Mahwah, NJ: Erlbaum.
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.
Robinson, B. F., & Mervis, C. B. (1998). Disentangling early language development: Modeling lexical and grammatical acquisition using an extension of a case-study methodology. *Developmental Psychology*, *34*, 363–375.
Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*, 283–331.
Rosa, E., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, *25*, 269–292.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). San Diego, CA: Academic Press.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
Simard, D., & Wong, W. (2001). Alertness, orientation, and detection: The conceptualization of attentional functions in SLA. *Studies in Second Language Acquisition*, *23*, 103–124.
Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277–303). Mahwah, NJ: Erlbaum.
Taraban, R., & Kempe, V. (1999). Gender processing in native and nonnative Russian speakers. *Applied Psycholinguistics*, *20*, 119–148.
Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203.
Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376–380.
Verfaellie, M., Koseff, P., & Alexander, M. P. (2000). Acquisition of novel semantic information in lesion location. *Neuropsychologia*, *38*, 484–492.
Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.

Williams, J. N. (2003). Inducing abstract linguistic representations: Human and connectionist learning of noun classes. In R. H. van Hout, A. Hulk, F. Kuiken, & R. J. Towell (Eds.), *The interface between syntax and the lexicon in second language acquisition* (pp. 151–174). Amsterdam: Benjamins.
Williams, J. N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Erlbaum.
Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, *53*, 67–121.

APPENDIX A

Table A1. Training items used in Experiments 1 and 2

Set 1 (Experiments 1 and 2)	Set 2 (Experiment 1)	Set 2 (Experiment 2)
I was terrified when I turned around and saw gi lion right behind me.	In the Safari park the children were excited when gi lions came up to the car.	In the Safari park the children were excited when gi lions came up to the car.
While I was digging in the garden gi birds came to eat worms next to me.	I was amazed when gi bird ate from my hand.	I was amazed when gi bird ate from my hand.
The children threw sticks at ul monkey in the tree.	The sound of ul monkeys in the rainforest was deafening.	The sound of ul monkeys in the rainforest was deafening.
The researchers studied ul bees from a safe distance.	The child screamed even though ul bee was on the other side of the nursery.	The child screamed even though ul bee was on the other side of the nursery.
I spent an hour polishing ro table before the dinner party.	The children all sat down at ro tables to play board games.	The children all sat down at ro tables to play board games.
The archaeologists worked for months to restore ro vases.	As I was passing I knocked over ro vase.	As I was passing I knocked over ro vase.
In the pub I asked my friend to get ne stool from the bar.	We hoped that ne stools would be delivered in time for the party.	We hoped that ne stools would be delivered in time for the party.
I could not sleep because of the chiming of ne clocks downstairs.	I looked up at ne clock on the church and realised that I was late.	I looked up at ne clock on the church and realised that I was late.
When I was out for a walk I patted gi dog and it bit me.	At bedtime I called to ul dog but he would not come.	My son was attacked by gi dogs as we walked down the path.

(continued)

Table A1. Continued

Set 1 (Experiments 1 and 2)	Set 2 (Experiment 1)	Set 2 (Experiment 2)
The farmer was kicked by gi cow when he tried to milk it.	The walkers could see that ul cow in the next field was in distress.	The walkers were surrounded by gi cows in the field.
Sitting under the tree I was bothered by gi flies.	We heard the sound of ul flies swarming around the dead animal.	The little boy tried for hours to kill gi fly in the kitchen.
The circus performer covered herself with gi snakes.	I was alarmed to see ul snakes approaching us very rapidly.	At the zoo none of the children wanted to touch gi snake.
I could hear ul mouse scurrying around in the roof.	While I was sitting in the kitchen gi mouse ran between my feet.	We were told that ul mice in the roof would breed if not killed quickly.
The fire brigade had to rescue ul cat from the top of the tree.	The old woman liked to sit for hours with gi cat on her lap.	In the night we heard ul cats fighting in the road outside.
I could hear ul pigs from the other side of the field.	The farmer could not stand the smell of gi pigs when he cleared out their pen.	I was distressed by the smell of ul pig coming from the farm next door.
I was terrified when I heard ul bears roaring in the distance.	The park warden reassured us that gi bears were tame enough to stroke.	We were able to observe ul bear wandering around the hillside across the valley.
After work I fell asleep on ro cushions in front of the TV.	When I had a backache I asked my wife to fetch ne cushions from the bedroom.	I told my son to put his head on ro cushion and go to sleep.
I spent the night on ro sofa and let my guests sleep in the beds.	While I was out for a walk I longed to sit on ne sofa in my living room.	The guests entered the living room and sat down on ro sofas.
I knocked over ro cup and the coffee spilled on my book.	I had to stretch to reach ne cup on the top shelf.	When the waitress dropped ro cups on the floor she was dismissed.
I can't move in my office because ro boxes are piled on the floor.	I asked the removal men to fetch ne boxes from the van.	I carefully packed my nephew's present in ro box before sending it to France.
I couldn't read the title of ne book that was on the top shelf.	In the library I laid my head on ro book and went to sleep.	I asked the librarian to fetch ne books from the stacks.
I had difficulty watching ne television when it was on the other side of the room.	The girl stayed up late watching ro television by her bed.	The scientist said that ne televisions in the office upstairs were interfering with his equipment.

(continued)

Table A1. Continued

Set 1 (Experiments 1 and 2)	Set 2 (Experiment 1)	Set 2 (Experiment 2)
In the gallery we all admired ne pictures from the other side of the room.	The art critics examined ro pictures for hours.	The office workers threw darts at ne picture of their manager.
At the fair they threw balls at ne plates.	The children broke ro plates during the meal.	From the kitchen I heard the sound of ne plate crashing to the floor.

APPENDIX B

TEST ITEMS USED IN EXPERIMENTS 1 AND 2: ACTUAL ORDER OF PRESENTATION (IN PHASE 1 OF THE TEST PHASE)

Original item / revised item (all participants in Experiment 2 received the revised items). Only the correct completion is provided.

Trained 1

The lady spent many hours sewing ro cushions.

The woman ran after ul mouse.

Generalization 1

The woman was stung by gi bees. / While lying in the field the woman was stung by gi bees.

The craftsman repaired ro stool. / While at the nursery the baby slipped off ro stool.

The cat wanted to eat ul birds. / The wolf wanted to eat ul birds.

The man spent the evening fixing ro clocks. / Before the battle the soldiers checked the time on ro clocks.

The waitress was told to go and clear ne table.

The boy held gi monkey. / The zookeeper struggled to control gi monkey.

The art collector went to Greece to collect ne vases. / The thieves searched the whole house for ne vases.

The hunters came to trap ul lion.

Trained 1 (continued)

After my meal I went to the sink to wash ro cup.
At the auction the farmer decided he wanted to buy ul pigs.
I searched everywhere for ne book.
The vet came to the farm to examine gi cow.
The tourists struggled to see ne pictures.
I was scared when I heard the hiss of gi snakes.
I left milk out for ul cat.
I packed all of my belongings into ro boxes.

Generalization 2

I heard the sound of gi bee. / While sitting by the wild flowers I heard the sound of gi bee.
The drunk fell over ro stools. / After the meeting we stacked up ro stools.
We enjoyed observing ul bird. / In the jungle the naturalist tried to take a picture of ul bird.
I adjusted ro clock. / I had to read the manual to find out how to adjust ro clock.
The customer pointed to his friends sitting at ne tables.
The boy gave bananas to gi monkeys. / The boy went into the forest to play with gi monkeys.
At the bottom of the drive they could see ne vase. / At the far end of the square the workmen fixed ne vase.
The game warden pointed to ul lions. / In the safari park the game warden pointed to ul lions.