An Investigation into the Role of Implicit Knowledge in Adult Second Language Acquisition

Anna Leonard Cook, M.A., M.Sc.

A thesis submitted in fulfilment of requirements for the degree of

to

University of Edinburgh

May 2009

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Anna Leonard Cook

Acknowledgements

I would like to express my deep gratitude to the many people who have contributed to this research. Most notably my supervisors Antonella Sorace, Martin Corley and Mitsuhiko Ota generously gave their time and expertise over the past years, and never failed to stretch and challenge me. Their guidance enhanced this project substantially, and their example inspired me to improve as a researcher. I owe them a particular debt of gratitude for their tremendous efforts in the later stages. It has been a pleasure to work with and learn from all of them.

I was lucky to spend three months working with Albert Costa, with whom I enjoyed many challenging and thought-provoking conversations. I can never thank him enough for the efforts that he made, and for the amount that I learnt from him. During the same time, Mireia Hernández, Iva Ivanova, Núria Sebastian-Gallès and Kristof Strijkers also provided insightful insights from which this thesis has benefited. In addition, Gerry Altmann, Axel Cleeremans, Caroline Heycock and Jan Hulstijn have generously offered me advice and guidance at various stages of the process.

I am very grateful to all the support staff at the University of Edinburgh. Katie Keltie and Toni Noble managed to find the answer to every admin question I generated, and they always did so with enthusiasm. The project could not have existed without Eddie Dubourg who taught me how to use E-Prime, or Morag Brown who rescued crucial files that had disappeared into the ether. Mike Bennett's patience and willingness to help fix every mistake I had made on the computer also deserves a particular mention.

Endless thanks are due to my friends and family, who supported me throughout the process and patiently put up with my endless absences. Specifically, my former office-mates Catherine Dickie, Evia Kainada, Katherine Messenger, Merilin

Miljan and Frances Wilson shared the good times and the bad, and generously provided chocolate whenever it was required. Alex Carter, Blair Duncan, David Fisher, Emily McCoombe, Arabella Millett, Charlotte Mooney and Hilary Wardle made themselves available whenever I was, and never complained when I had to abandon them. Bob Cann deserves a special mention for knowing enough to talk me out of it but choosing not to, as does Vida Sherkat-Khameneh for letting me treat her home like a restaurant when working late, even though I was supposed to be looking after her. My eternal gratitude goes to Chris Cann, Keith Leonard, Inge Cryer and Phyllis Leonard who gave me the confidence, ambition and motivation to begin and to finish the process. I wish Inge was able to share this day. Finally, I could not have completed the project on time without Sara Kheradmand, who never complained when I stole the laptop. For every lift home, dinner made, and well-timed glass of wine, thank you. But most of all, thank you for the inspiration.

This research was supported by studentship number PTA-030-2003-00487, awarded by the Economic and Social Research Council.

Abstract

This thesis investigates the role of implicit knowledge in second language acquisition, presenting five experiments and related simulations based on artificial grammar learning. It examines whether second language learners can acquire implicit knowledge of noun–verb agreement. In addition it tests whether their ability to do so is influenced by the number of words that intervene between the relevant noun and the finite verb in the input sentences, as this affects performance in artificial grammar learning, the serial response time task and the statistical learning paradigm.

Experimental participants were exposed to a modified version of Persian or Basque while performing a memory task. Next, two grammaticality judgement tests (one timed and one untimed) and a sentence correction task assessed whether the participants had acquired either the target noun–verb agreement or pairs of adjacent words that appeared frequently in the learning items. Performance reflecting implicit and explicit knowledge was distinguished according to three criteria based on R. Ellis (2005) and according to the assumption that the former is not under conscious control.

Participants' performance suggested that they had implicit knowledge of the adjacent word pairs. Similarly, the results indicated that they had implicit knowledge of subject–verb agreement when a single word intervened between the subject and the verb, but not with two intervening words. Connectionist simulations supported the results and indicated that performance was unlikely to improve if more exposure were given.

Although the influence of additional factors is discussed, the results generally supported the view that an increase in the number of intervening words reduces learning outcome. The intriguing similarity between the results of this thesis

and previous research in artificial grammar learning, the serial response time task and statistical learning experiments suggests that future research should directly compare the paradigms to ascertain whether similar learning processes are engaged in each case.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Introduction

Exposure is widely recognised to be a vital element in successful second language acquisition regardless of whether it occurs through deliberate hypothesis testing (DeKeyser, 1995), parameter resetting (White, 2003) or frequency tallying (N. C. Ellis, 2003). However, input alone is not sufficient for native–like proficiency to develop. Firstly, performance tends to be higher when learners receive metalinguistic instruction in addition to exposure (Erlam, 2003; Klapper & Rees, 2003; Norris & Ortega, 2000). Secondly, adult learners in immersion situations rarely achieve native–like proficiency, in some cases despite many years' residence in a target–language speaking country (Johnson & Newport, 1989; Lardiere, 1998; Long, 2003). Even the few that reach sufficiently high standards to be classified as near–native still tend to diverge from true native–like performance with structures at the syntax–discourse interface including tense and aspect distinctions (Coppieters, 1987) and null subjects (Belletti, Bennati, & Sorace, 2007). Overall therefore, although exposure undoubtedly increases performance in a second language, there remain limits on what can be acquired.

This thesis investigates the limits on second language acquisition as it occurs during exposure and when no formal grammar instruction is provided. It focuses on one form of knowledge that may be acquired under such circumstances: implicit knowledge of which the learner is not aware. (Section 2.1 will describe this phenomenon in more detail.) We ask whether implicit knowledge of second language structures can be acquired and whether restrictions apply.

Theoretically, the role of implicit knowledge could be limited in three ways. Firstly, situational variables including the purpose of an interaction, the experimental task or the quantity of input may limit acquisition. Secondly, learners may acquire accurate implicit knowledge of the target language, which they are unable to express fully in their performance. Thirdly, learners may only be able to acquire accurate implicit knowledge of a subset of the target language structures. This thesis investigates the third option.

Researchers investigating the acquisition of implicit knowledge in second language acquisition have studied many individual target structures including article use (E. Hauser, 2000; Seliger, 1979), verb complements (Han & Ellis, 1998), subject–verb inversion (Hulstijn & Hulstijn, 1984) and animacy (Williams, 2005). However there have been few attempts to create a general framework for determining what is learnable. We believe that such a line of enquiry would be more valuable as well as efficient, providing the context in which to interpret the results of individual experiments.

There has been a sustained effort to build such a general framework in the artificial grammar learning paradigm, from which the study of implicit knowledge in second language acquisition originally developed (A. S. Reber, 1967, 1989). Many investigations have been focused on two questions, both of which are relevant to language: whether learners acquire implicit knowledge of contingencies between non-adjacent elements (Johnstone & Shanks, 2001; Perruchet & Pacteau, 1990), and whether they acquire abstract categories as well as surface patterns (Gómez & Gerken, 2000; A. S. Reber, 1969). This thesis focuses on the first issue and asks whether second language learners can acquire implicit knowledge of agreement relations affecting non-adjacent words. Thus it targets a particular dimension along which certain target structures may vary.

Specifically, the experiments were designed to investigate whether learners could acquire implicit knowledge of subject–verb and object–verb agreement. The number of words intervening between the relevant noun and the verb was manipulated between experiments in order to test the effects of different contingency lengths. A series of connectionist simulations was then conducted to model the human data and generate hypotheses for future research. Thus, the experimental and simulation results provided evidence about whether implicit knowledge of linguistic structures containing longer–distance contingencies is learnable.

1.2 Structure of the Thesis

The thesis is structured as follows. Chapter 2 defines the terms *implicit knowledge* and *implicit learning*, and describes the form of implicit knowledge that is the focus of the current thesis. The bulk of the chapter then reviews previous experimental research into implicit learning and the acquisition of implicit knowledge, focusing on second language acquisition but also drawing on experiments in the artificial grammar learning, serial response time task and statistical learning paradigms. The findings are shown to be consistent with the suggestion that contingencies between local elements are acquired but that those between distant elements are not acquired. Finally, Section 2.8 specifies the research questions that are addressed in this thesis: how to distinguish between performance based on implicit and explicit knowledge, whether relevant implicit knowledge would be acquired, and whether the acquisition of implicit knowledge would be restricted to shorter contingencies.

Chapter 3 first provides an overview of the experimental design. It then contains a discussion of a number of relevant methodological issues. Specifically it considers the amount of exposure to the target language that experimental participants need to receive, how best to differentiate implicit and explicit knowledge, and how to measure the length of an agreement dependency in second language sentences.

This thesis consists of five individual experiments, each presented in a separate chapter including its own objectives, method and results. Chapter 4 reports Experiment 1, a conceptual replication of Johnstone and Shanks (2001) in which participants were exposed to the output of an artificial biconditional grammar. In the remaining experiments participants were exposed to input in a novel foreign language. Experiments 2 and 3 (reported in Chapters 5 and 6 respectively) used subject–verb agreement in Persian as the target structure. Chapters 7 and 8 then presented Experiments 4 and 5, in which the target structures were object–verb and subject–verb agreement in Basque respectively.

In addition to Chapters 4 to 8, there are two further data–driven chapters. The first, Chapter 9, reports analyses in which data from the different experiments were combined. First, learning phase performance was compared across the experiments. Then, reaction times and the relation between participants' previous language learning experience and their experimental performance were both examined. The second, Chapter 10, reports a series of connectionist simulations that aimed to replicate and extend the experimental results to demonstrate what can be acquired from the statistical distribution of the input.

Chapter 11 is devoted to a discussion of the study's findings. First, the results are considered in terms of the three research questions previously developed in Chapter 3. Next, it contains a discussion of whether the conclusions can be generalised to second language acquisition outside of the laboratory, focusing on the issues of frequency, abstraction, attention and phrase structure. Finally, Chapter 12 presents concluding remarks and makes suggestions for further research.

CHAPTER 2

Previous Research into Implicit Knowledge

Adults learning a second language have a variety of mutually compatible strategies at their disposal, all of which may contribute to performance in varying degrees. Many learners take classes in which the target language and the rules of the grammar are taught explicitly using metalinguistic descriptions. In addition, they may spontaneously notice new regularities in the input, and then deliberately use a problem–solving or hypothesis–testing approach in an attempt to discover the relevant rule. One difference between these two strategies is that the former relies largely on metalinguistic input whereas the latter only requires exposure to the target language in use. Nevertheless, in both cases learning is deliberate and accompanied by awareness (DeKeyser, 1995).

Other forms of second language acquisition (SLA) may occur without the learner being aware of the process. In general therefore they can be defined as implicit. Generative approaches to the study of SLA are primarily theories of interlanguage representation rather than of development (White, 2003). Nevertheless, if learners are able to reset the parameters of universal grammar following exposure to second language (L2) input (as claimed by Kanno, 1997; Schwartz & Sprouse, 1996), the process is likely to be implicit and to result in implicit knowledge. Learners without linguistic training generally have no awareness of these parameters.

Although parameter resetting would be implicit, the terms *implicit learning* and *implicit knowledge* are typically used to describe a non-language-specific learning process in which sensitivity to a given target domain is gradually acquired (N. C. Ellis, 2003). This form of SLA was the focus of the research presented in this thesis. Therefore the following section defines and describes the phenomenon in more detail.

The remainder of the chapter is then organised as follows. Section 2.2 will describe one factor that might limit the acquisition of implicit knowledge: the length of the target contingency. Previous research in implicit learning and the acquisition of implicit knowledge in second language acquisition is discussed in Section 2.3 with respect to this feature. Sections 2.4, 2.5 and 2.6 then do the same for research in the artificial grammar learning, serial response time, and statistical learning paradigms respectively. Finally, three research questions are developed in Section 2.8.

2.1 Definition and Characteristics of Implicit Knowledge

2.1.1 Definitions

This section begins the discussion of implicit knowledge in second language acquisition by defining the term *implicit knowledge*. It examines both what is meant by *knowledge*, and the meaning of the term *implicit*.

For the purposes of this thesis, *knowledge* is defined simply as the outcome of learning. Thus, it could be declarative knowledge of a fact or procedural knowledge of an action, depending on the type of learning involved (Berry & Dienes, 1993; DeKeyser, 2003). *Learning* on the other hand is the process that creates sensitivity to the structure of a domain to allow generalisations to novel stimuli (Cleeremans, Destrebecqz, & Boyer, 1998). Note that in this thesis the terms *learning* and *acquisition* are used synonymously.

The second stage of defining implicit knowledge is considering the meaning of *implicit*. By definition, the term refers to a lack of awareness (Gómez, 1997; Green & Shanks, 1993; Hulstijn, 2002a; Manza & Bornstein, 1995; A. S. Reber, 1989). (Note also that *consciousness* and *awareness* will be treated as synonyms.) The opposite term, *explicit*, therefore refers to the presence of awareness. Thus, *implicit learning* is the acquisition of structural sensitivity without awareness of the learning process, while *implicit knowledge* is structural sensitivity of which the learner is not aware (Berry & Dienes, 1993; DeKeyser, 2003).

It is still an open empirical question whether implicit knowledge must result from implicit learning. Although implicit and explicit knowledge are likely to derive from different learning processes, awareness may not be the dimension on which these processes differ. It is equally possible that implicit knowledge stems from data–driven processes and explicit knowledge from conceptually driven learning modes. Therefore it is important to specify that this study is investigating implicit knowledge rather than implicit learning (Berry & Dienes, 1993).

2.1.2 Characteristics of Implicit Knowledge

Researchers have theorised that awareness exists to increase control over behaviour (Cleeremans & Jiménez, 2002; Frensch & Miner, 1994; Maia & Cleeremans, 2005). Specifically, Cleeremans and Jiménez described a tripartite system in which implicit knowledge has small effects on behaviour that cannot be controlled. On the other hand, explicit knowledge is available for conscious control. It is believed to have larger effects, and therefore control is necessary to ensure that the knowledge is only utilised appropriately. Finally, there is a third type, automatised knowledge. This is a sub-type of explicit knowledge that has been practised extensively until it can be used quickly and automatically (DeKeyser, 2001). Therefore it also has large effects, but as it has been sufficiently tested over time it no longer requires conscious control.

Nevertheless, the control theory does not have unanimous support (e.g. Shanks, Rowland, & Ranger, 2005), and it is not the only theory that has been proposed to account for consciousness. For example, Dienes and Perner (1999) suggested that consciousness requires higher order thought. That is, for knowledge to be explicit a learner must know that they know it. In contrast, O'Brien and Opie (1999) offered a neurological account, asserting that stable activation patterns produce awareness. Debating the merits of these and other proposals is beyond the role of this thesis. Instead, relevant reviews can be found in Atkinson, Thomas, and Cleeremans (2000), Dehaene and Naccache (2001), Dennett (2001) and Frith, Perry, and Lumer (1999).

The control theory will be adopted in this thesis. Rather than researching implicit knowledge in general, the experiments investigate a type of implicit knowledge that cannot be controlled. Under this assumption, if a learner had uncontrollable implicit knowledge of a linguistic structure, they would automatically sense the presence of an error whenever they encountered an ungrammatical exemplar. Nevertheless they would still be able to decide consciously how to react to that ungrammaticality, by ignoring it, providing a recast, or answering *incorrect* in a grammaticality judgement test for example. Logically, if a learner does not control the use of implicit knowledge, it must be triggered by the input or the situation. For example, implicit knowledge of object–verb agreement would be accessed whenever transitive sentences were encountered but not otherwise. Thus, access to implicit knowledge is inflexible because it would be obligatory in contexts where the appropriate input is provided and impossible otherwise. Consistent claims have been made by Berry and Dienes (1993), N. C. Ellis (2005), Hulstijn (2002a), Meulemans and Linden (2002) and Squire, Hamann, and Knowlton (1994).

2.1.3 Acquiring Implicit Knowledge

As stated in Section 2.1.1, implicit knowledge may not always derive from implicit learning. Nevertheless implicit and explicit knowledge are likely to be acquired via different processes. Thus, this section considers how implicit knowledge is acquired.

The acquisition of implicit knowledge has been described as a tallying process (N. C. Ellis, 2002a, 2005). This theory claims that the link between two representations is strengthened when they are active simultaneously (Frensch & Miner, 1994). Note that Hebb (2002) also made similar claims although he did not relate them specifically to *implicit* knowledge. Structural priming experiments also support the proposal (Ferreira & Bock, 2006). Priming effects following the presentation of a particular structure have been detected several minutes after presentation of the prime (Bock & Griffen, 2000; Bock, Dell, Chang, & Onishi, 2007; Boyland & Anderson, 1998; Ferreira, Bock, Wilson, & Cohen, 2008). Therefore the phenomenon is likely to result at least in part from changes in the strengths of neural representations. Overall, the tallying theory implies that implicit knowledge is not only used automatically, but also acquired automatically (cf. Gómez, 1997).

McClelland, McNaughton, and O'Reilly (1995) and McClelland (1998) also provided evidence that two learning modes are employed in the human brain. Based on studies of amnesia they argued that memories are originally stored explicitly as individual instances in the hippocampus. As multiple exemplars are repeatedly presented, the neocortex gradually becomes sensitive to the structure of the domain and develops a different knowledge store that may be implicit. This proposal implies that implicit knowledge is acquired more slowly than explicit knowledge. This is consistent with other claims made by DeKeyser (2003), N. C. Ellis (1993) and Norris and Ortega (2000) specifically for SLA.

Finally, note that the relationship between implicit and explicit knowledge is not addressed in the current project (for which see N. C. Ellis, 2005; R. Ellis, 1994). However it is assumed that learners may acquire both simultaneously. Thus, the presence of the one does not entail the absence of the other. Nevertheless as stated above, explicit knowledge is likely to have larger effects on behaviour than implicit knowledge (Cleeremans & Jiménez, 2002). Therefore the latter is easier to detect in the absence of the former.

2.1.4 Section Summary

To summarise, implicit knowledge is not available to consciousness by definition. In addition, the type of implicit knowledge investigated in this thesis is not subject to conscious control. It is both acquired and accessed automatically as a consequence of a tallying process, and it takes longer to acquire than does explicit knowledge. Note that alternative forms of implicit knowledge may exist, but they are unlikely to be detected by the technique employed in the current experiments.

2.2 Restrictions on Implicit Knowledge

Section 2.1 described a form of implicit knowledge that is acquired continuously (e.g. N. C. Ellis, 2002a). That is, the brain gradually acquires sensitivity to the structure of a given target domain whenever input is received. Nevertheless, it may not be possible to acquire all linguistic structures in this way. This section considers the type of structure that learners may be able to acquire. This is important because there could be substantial differences between structures. If an experiment investigated the acquisition of subject–verb agreement, any test would be designed to detect only implicit knowledge of that structure. However participants may have acquired implicit knowledge of a different pattern in the input, such as the frequencies of the individual words. In such circumstances the participants' knowledge would not be detected by the test and researchers may claim that no such learning had occurred. However, when an experiment does not provide any evidence of implicit knowledge, the results only imply that the specific target structures were not acquired. (Note that this claim is analogous to the Information Criterion proposed by Shanks and St. John (1994) for explicit knowledge.) This section therefore considers which regularities may be amenable to implicit learning and the acquisition of implicit knowledge.

Krashen (1982) and A. S. Reber (1993) originally claimed that implicit learning and knowledge were well suited to complex domains whose rules would be too difficult to induce explicitly in SLA and artificial grammar learning (AGL) respectively. Robinson (1996) tested this claim in second language acquisition. He gave learners of English as a second language additional training in two rules: how to form pseudoclefts of location (e.g. *Where John and Mary live is in Chicago not in New York*), and the use of subject–verb inversion when adverbials of movement or location are fronted (*Into the house ran John* versus *Into the house John ran*). Following the intuitions of language teachers, the former was described as a hard rule and the latter as an easy rule (note that the participants were already intermediate level learners of English). Robinson hypothesised that, for the harder rule, participants in an incidental condition designed to encourage implicit learning would outperform learners in an explicit inductive condition. However, as there were no significant differences between the groups the results did not support the prediction. Thus, difficulty did not influence whether a structure could be acquired implicitly.

Robinson (1996) considered whether a rule was simple or complex based on the intuitions of language teachers, who would have been experienced in teaching explicitly. Therefore their categorisations would be based on the difficulty of learning a structure in that mode. However there may be independent scales of difficulty for implicit and explicit knowledge and learning (R. Ellis, 2006). If this is the case, it would be impossible to determine the structures for which implicit knowledge can be acquired, based on whether or not explicit knowledge can be acquired. Instead, the possibility of acquiring implicit knowledge must be assessed against the difficulty of acquiring *implicit* knowledge. According to R. Ellis, frequency, saliency, regularity and functional complexity all affect the difficulty of a structure for implicit learning. Explicit learning on the other hand is mainly affected by the ease with which the target regularity can be described as a rule.

As stated above, implicit knowledge is hypothesised to be an automatic consequence of processing multiple items simultaneously (Frensch & Miner, 1994). Under this proposal, when two neural representations are activated at the same time the link between them is strengthened (Hebb, 2002). Clearly, if two elements are separated by a large amount of intervening material, processing of one is likely to have been completed before the other is encountered unless the task requires otherwise. The exact amount of intervening material required is unclear. Nevertheless, under this theory, relevant implicit knowledge of a longer–distance dependency is less likely to be acquired than implicit knowledge of a local contingency.

Researchers examining the acquisition of longer structures have studied two versions that are both relevant to language: long–distance dependencies (also termed long–distance contingencies) and larger chunks. Dependencies will be defined as rules that specify the form of one element in a sequence based on the form of another element. In local versions these two elements are adjacent, while in long–distance versions other material intervenes. The specific length of the structure is calculated based on the number of steps between the two relevant elements. That is, the subject–verb agreement in (1*a*) is a one–step dependency because there is a single step between *John* and *walks*. In (1*b*) on the other hand, it is a two–step dependency because there are two steps, from *John* to *always* and from *always* to *walks*. Finally, in (1*c*) it is a seven–step contingency.

(1) a) John walks.
b) John always walks.
c) John, the man I saw yesterday, always walks home.

In long–distance dependencies the identity of the intervening material is irrelevant. This is not the case with chunks. In a bigram, or two–item chunk, the identity of the first element determines that of the second[1]. Thus, it is indistinguishable from a one–step contingency. In a trigram however, the identity of the final element depends on the *combination* of the first two and in a four–item chunk it is determined by the first three. (Clearly intermediate forms may also exist, whereby the final element in a four–item chunk depends on the combination of the first two but the third is in free variation. However, this option is not discussed because we are not aware of any previous research into this type of structure.)

Long–distance dependencies and chunks both occur in language. A common example of the former is agreement, as in the examples given in (1). On the other hand the latter type is more likely to be a formulaic sequence of lexical items such as the trigram in (2). The current thesis focuses on the acquisition of long–distance dependencies rather than of chunks. Nevertheless, as there is only a limited

[1]Strictly speaking, the term *bigram* refers to a pair of adjacent letters, but following Robinson (2005) we will use it to denote any pair of adjacent elements in a sequence including neighbouring words in linguistic input.

amount of previous research into each type, both are included in the reviews that follow. Specifically, Gómez (2002), Newport and Aslin (2004) and Peña, Bonatti, Nespor, and Mehler (2002) have considered the acquisition of two–step dependencies in the statistical learning paradigm while Johnstone and Shanks (2001) and Mathews et al. (1989) studied four–step versions in artificial grammar learning. (These experimental paradigms will be described in the sections that follow.) We are not aware of previous research into other combinations of dependency length and paradigm, which would be necessary to identify an upper limit above which implicit knowledge cannot be acquired.

(2) Beanz meanz Heinz.

In contrast, chunk length has been manipulated more thoroughly within a single paradigm. For example, Curran and Keele (1993), Jiménez, Méndez, and Cleeremans (1996) and Shanks et al. (2005) investigated the acquisition of trigrams in the serial response time task, while Cleeremans and McClelland (1991) also included four– and five–item chunks. However before discussing research based on chunks, it is necessary to consider the relationship between performance with them and with long–distance dependencies. Jiménez et al. reported that performance with three–step dependencies was lower than that with chunks of the same length. Thus, if participants learn a chunk, it would not necessarily follow that they can also acquire a contingency of the same length. However, should they fail to develop implicit knowledge of a chunk, then they would be unlikely to learn the corresponding long–distance dependency. Overall therefore, experiments focusing on chunks can provide evidence relevant to long–distance dependencies, although performance with the two types of structure is unlikely to be identical.

In summary, there is reason to believe that the length of the target structure may impact upon the acquisition of relevant implicit knowledge. Therefore the next section discusses research in second language acquisition in the context of dependency length.

2.3 Implicit Knowledge and Implicit Learning in Second Language Acquisition

This section reports experimental studies that allow the degree of control necessary to differentiate implicit from explicit knowledge and learning. Although the focus of the current thesis is on implicit knowledge, the review covers experiments in both implicit knowledge and implicit learning. As discussed above in Section 2.1, many researchers maintain that the two phenomena are intrinsically linked. Including both therefore increases the range of available evidence. Nevertheless they are separated into different subsections for clarity. Each subsection begins with a discussion of how implicit knowledge or learning is usually detected. The length of the target structure(s) is also specified where applicable.

2.3.1 Implicit Knowledge

Different tests are often employed to detect implicit and explicit knowledge respectively. Often experimenters use spoken language production tasks to detect the former (e.g. Erlam, 2006; E. Hauser, 2000; Hulstijn & Hulstijn, 1984). Implicit knowledge is likely to contribute to performance on these tasks, because the focus is normally on meaning rather than on form, and because the tasks tend to incorporate time constraints that restrict the use of explicit knowledge (Han & Ellis, 1998; Krashen, 1982). Nevertheless, although the use of explicit knowledge is restricted, it may not be completely eliminated.

On the other hand explicit knowledge can be assessed by asking participants to report the rules of which they are aware. These verbal reports are often compared to grammaticality judgement tests or production tasks. If the other tests reveal knowledge of regularities that were not included in the verbal report, such patterns are assumed to be known implicitly. The type of implicit knowledge described in Section 2.1 should not be available for verbal report, because they are produced in the absence of directly triggering input. Therefore performance on the task is more likely to be the result of explicit knowledge. However there is a problem with this approach because it relies on the assumption that participants include all relevant explicit knowledge in a verbal report (Shanks & St. John, 1994). This may not actually be the case, and a learner may report only a subset of their explicit knowledge. As a result, the unreported explicit knowledge would be mistakenly classified as implicit. Nevertheless, studies relying on verbal report are included in this review simply because the existing literature does not provide sufficient alternative sources of evidence.

More recently, R. Ellis and colleagues investigated which tests encourage the use of implicit and which of explicit knowledge (Han & Ellis, 1998; R. Ellis, 2005). They discovered that tests including a time constraint were more likely to be

		Distance	
		Near	**Far**
Animacy	**Animate**	gi	ul
	Inanimate	ro	ne

Table 2.1: Classifiers from the artificial language used in Williams (2005). They are divided as to whether their referents must be near or far, and animate or inanimate.

completed using implicit knowledge than explicit knowledge. This included timed grammaticality judgement tests, an oral production test and an imitation test. In the latter, participants heard incorrect exemplars, which they are asked to repeat from memory *"in correct English"*. If the items were spontaneously corrected, the learners were assumed to know the rule that had been violated (R. Ellis, 2005).

The participants' performance on the different tasks was subjected to a factor analysis that revealed two separate factors contributing to performance. The oral tasks and the timed grammaticality judgements loaded strongly onto the first, and the untimed grammaticality judgements (particularly the ungrammatical items) and the metalinguistic test onto the second. Thus, the authors concluded that those tests including a time constraint mainly detected implicit knowledge and those without the time constraints allowed the use of more explicit knowledge. These findings are invaluable when designing valid tests for use in future experiments, allowing researchers to improve on the use of verbal report. We now begin to discuss the findings of previous research, in each case specifying the means by which implicit and explicit knowledge were distinguished.

Seliger (1979) conducted an early investigation into the role of implicit knowledge in the acquisition of form–form correspondences in a second language. He focused on use of the indefinite article, examining participants' use of *a* versus *an*. This is a one–step dependency because the selection of the correct indefinite article depends purely on the immediately following phoneme. Learners first completed a picture–naming task, producing the indefinite article followed by the relevant noun. Next, they were asked to verbalise the rule governing their performance. Four of the fifteen were able to report a rule, but in three cases this was not related to their earlier productions. At least for these participants therefore, explicit knowledge as expressed through verbalised rules was unlikely to have been the main origin of their language production. Instead they may have been using implicit knowledge.

Whereas many of the experiments reported so far investigated implicit knowledge of syntactic features and form–form mappings, Williams (2005) studied form–meaning relations. Participants were exposed to English sentences with the addition of the classifiers shown in Table 2.1. They were informed that *gi* and *ro* indicated nearby referents, and *ul* and *ne* more distant referents. However unbeknownst to the participants there was also an animacy distinction. *Gi* and *ul* were used for animate nouns, and *ro* and *ne* for inanimate ones. Following training during which their attention was focused on the distance dimension, the participants' verbal reports indicated that eight of them also became aware of the animacy pattern. The remaining participants, assumed not to have explicit knowledge, were nevertheless able to select the appropriate classifier at a level significantly above chance. Presumably therefore this was a result of implicit knowledge.

The oral imitation test developed by R. Ellis (2005) was used by Erlam (2006). Second language speakers were asked to repeat ungrammatical sentences *"in correct English"*, without being informed of the presence of the errors. They spontaneously corrected 35%, a result which was interpreted as evidence of implicit knowledge. It is unclear which specific rules they had acquired however, because the test included a wide variety of structures.

In contrast, Hulstijn and Hulstijn (1984) provided mixed evidence about the presence of implicit knowledge. They tested adult learners of Dutch on their production and verbally reportable explicit knowledge of two structures: subject–verb inversion and verb–final word order in subordinate clauses. The dependency length could not be specified in either case, because the learners had been exposed to the language prior to the experiment, and their input had probably included exemplars with different dependency lengths. In the production task the participants who reported explicit knowledge of the verb final construction outperformed those who either reported an erroneous rule or no rule at all. Therefore, at least some of their production was likely to have resulted from explicit knowledge. However a different pattern of results was reported for inversion. Both groups' production was approximately 80% accurate regardless of whether or not they could report the rule verbally. Thus, explicit knowledge did not appear to influence production that was instead likely to have been the result of implicit knowledge. In summary, the participants appeared to have implicit knowledge of inversion but not of verb–final word order in Dutch.

R. Ellis (2005) also reported ambiguous findings. He considered seventeen English target structures including the regular past tense, the indefinite article, question tags, and adverb placement. The participants' mean timed test accuracy was 54% (95% confidence interval: 51.6% to 56.4%). As the confidence interval did not include the 50% chance level, performance may have been significantly above chance, providing evidence of minimal implicit knowledge. However the data were not divided between the seventeen target structures, so it is impossible to know which structure(s) (and which dependency lengths) the learners knew implicitly.

Not all experiments have revealed evidence of implicit knowledge. E. Hauser (2000) tested learners of English on definite article use in place names (also a one–step dependency between adjacent words). Those who reported awareness of the rules improved the accuracy of their production, but those without conscious awareness did not. Thus, their explicit knowledge alone was sufficient to account for the former group's performance.

Han and Ellis (1998) examined the participants' knowledge of the complements that can follow different verbs, a one–step dependency. They did not compare the participants' performance to that of naive controls or even to chance[2]. Nevertheless the authors reported that mean accuracy on the timed grammaticality judgement tests was 42% against a chance level of 50%. Therefore the experiment did not provide any evidence that the participants had implicit knowledge of the correct verb complements.

In summary, the empirical evidence suggests that learners often have little or no implicit knowledge of the relevant target patterns. Nevertheless they can sometimes acquire a degree of implicit knowledge of second language structures, including the choice of *a* versus *an* in English, Dutch subject–verb inversion, and the role of animacy in classifier selection in an artificial language. However, these results depended on verbal report, which may have overestimated the role of implicit knowledge. Therefore it is important to investigate the findings further using a different methodology for distinguishing between implicit and explicit knowledge.

[2]Untrained controls often do not actually perform at chance levels (e.g. Michas & Berry, 1994; Saffran, 2001). Therefore, it is advisable to employ such participants to determine how someone would respond if they had no knowledge of the target structure, rather than relying on statistical approximations.

2.3.2 *Implicit Learning*

A large proportion of research into implicit second language acquisition focuses on the question of whether implicit or explicit learning tasks lead to superior performance, comparing participants in a condition designed to engender implicit learning with those in an explicit condition. The results usually indicate that the implicit group under–performs their explicitly–trained counterparts (Norris & Ortega, 2000). However as the group's performance is often not compared to untrained controls or even to chance, it is uncertain whether they actually acquire any aspects of the target language.

Implicit learning was defined in Section 2.1 as occurring without awareness of the learning process, and explicit learning as occurring in the presence of such awareness. However, the experimental tasks designed to encourage either learning mode tend to manipulate intention rather than awareness. As such, tasks during which implicit learning is expected tend to be tasks that do not require focus on the relevant aspects of the grammar. They are either communicative (e.g. Morgan-Short, 2007; Robinson, 1996), or based on memorisation (e.g. N. C. Ellis & Schmidt, 1997; Robinson, 1996, 1997a, 1997b). Explicitly–biased tasks can be divided into two categories. One type encourages an inductive or rule–search approach by including instructions to search for rules or input enhancements that highlight the relevant grammatical features (e.g. Robinson, 1996). On the other hand participants in explicit deductive or instructed conditions are taught the target rules, usually in conjunction with exposure to input (e.g. N. C. Ellis, 1993; Morgan-Short, 2007; Robinson, 1996).

When participants are deliberately trying to learn a particular target structure, they are aware that they are doing so and therefore they must be learning explicitly. However when they are not deliberately attempting to learn, it is conceivable that they may notice salient regularities explicitly. Thus, this procedure risks mistakenly categorising explicit learning as implicit learning. Nevertheless the influence of explicit learning is likely to be larger for those in intentional conditions. If incidentally trained participants outperform intentionally trained ones therefore, some degree of implicit learning is likely to have contributed to their performance Reber1989.

Participants in the first pair of studies to be reported were trained in an incidental condition expected to engender implicit learning. However there were neither any untrained controls nor any learners given intentional instructions to explore the consequences of explicit learning. First, N. C. Ellis and Schmidt (1997) told learners to memorise forty input sentences in an artificial language (half of these were presented twice). Subsequently, their knowledge of adjective–noun agreement and subject–verb agreement in both intransitive and transitive sentences was assessed. The first two were one–step dependencies, but the final one was a two–step dependency because the object noun intervened between the subject and the verb in the transitive item — see (3). The test was a grammaticality judgement. As this test did not include a time constraint, explicit knowledge may have been used. (Remember that we do not assume that implicit learning always results in implicit knowledge.) The participants' mean accuracy was 49% (95% confidence interval: 42% to 56%) for adjective–noun agreement, 57% (49% to 65%) for intransitive subject–verb agreement and 50% (43% to 57%) for transitive subject–verb agreement. In all cases therefore the 50% chance level was within the 95% confidence interval. In the absence of inferential statistics, there is no evidence that the participants had acquired any of the target structures in incidental conditions.

(3) a) Adjective_{sg} — Subject_{sg} — Verb_{sg}
b) Subject_{sg} — Verb_{sg}
c) Subject_{sg} — Object_{pl} — Verb_{sg}

Similarly Robinson (2005) trained participants in incidental conditions during which they were encouraged to process the input for meaning, a more naturalistic task than memorisation as used by N. C. Ellis and Schmidt (1997). A subsequent untimed grammaticality judgement test was used to examine whether the learners had acquired three Samoan target rules: marking the locative by placing an *e* particle directly before the subject of a transitive verb, placing an *i* particle immediately before an adverbial phrase, and optionally incorporating direct object nouns into the verbs. Examples of the structures are shown in (4), (5) and (6). The ergative and locative structures were a type of one–step dependency, but between the form of the particle and an abstract category rather than between two surface elements. While the participants did not demonstrate any significant knowledge of the ergative markers or incorporation, they performed at above chance levels with locative items. In contrast to N. C. Ellis and Schmidt therefore, Robinson provided evidence that implicit learning could succeed in second language acquisition, at least with some target structures. The differences between the two experiments may be the result of the respective learning tasks.

(4) Ergative: Ave e le tama le taavale.

		Drive	ergative the	boy	the	car.	
		"The boy drives the car."					
(5)	Incorporation:	Inu-pia	le	tama.			
		Drink-beer	the	boy.			
		"The boy drinks the beer."					
(6)	Locative:	Malaga	le	teine	i	le	vaa.
		Travel	the	girl	locative	the	boat.
		"The girl travels in the boat."					

Next, experiments contrasting implicit and explicit learning conditions will be reported. First, participants in an experiment by N. C. Ellis (1993) were trained on soft mutation in Welsh, in which the initial consonants of certain words are altered depending on the phonological context — see the example in (7). This is a one–step dependency, as the mutation is triggered by the identity of the following consonant. Some participants only saw examples while others also received metalinguistic information. The results of a series of grammaticality judgement tests suggested that the former group had acquired significantly less knowledge about the target structure than the latter. However the implicit participants' performance did improve slowly during the experiment as more exposure was provided, suggesting that some degree of learning took place.

(7)	a)		Caerdydd
			Cardiff
	b)	i	Gaerdydd
		To	Cardiff

Similarly, DeKeyser (1995) trained participants on an artificial language (named Implexan) with one group given metalinguistic instruction and the other asked to memorise the input. The target structure was subject–verb agreement, a one–step contingency in the sentences used in the experimental materials. (The experiment was also designed to assess allophone selection, but insufficient tokens were obtained to permit statistical analyses.) Production tasks were used to assess whether the learners had acquired the target structure. As in the other studies, the explicitly trained group significantly outperformed their incidental counterparts with the agreement structure. In fact, there was no evidence that the latter had acquired the target structure, as their mean accuracy was 49.45% compared to a chance level of 50%. Thus, the experiment did not provide any evidence of implicit learning.

Robinson (1996) compared learning in four rather than two conditions. Second language learners of English were given additional training on the two target rules described in Section 2.2 (pseudoclefts of location and subject–verb inversion). Implicit and incidental groups were not focused on the rule and were therefore assumed to be learning implicitly. The former was asked to memorise the input, while the latter was focused on the meaning. A rule search group was told to induce the rule, while an instructed group was taught the rule explicitly. The results of a subsequent grammaticality judgement test confirmed that the final group was the most accurate. Nevertheless the implicit and incidental groups may also have acquired the two English target structures to some degree. Their judgement accuracy was always more than 55%, reaching 70% in the case of the incidental learners with the easier rule (no measure of dispersion was reported). However, in the absence of a comparison group of untrained control participants it is impossible to confirm conclusively that the group had acquired the target structure. As the incidentally trained participants marginally outperformed those in memorisation conditions, the results also suggested that processing for meaning increases learning outcomes. This could partially account for the contrast between N. C. Ellis and Schmidt (1997) who did not report any significant implicit learning in a memorisation task, and Robinson (2005) who did following a meaning–orientated learning task.

The remaining studies reported in this section contrasted implicit learners against a comparison group of participants either given no exposure or trained on similar but non-rule governed input. Firstly, Michas and Berry (1994) exposed learners to stimuli written in the Greek alphabet and trained them on the grapheme–phoneme correspondences. As this was not a form–form contingency, dependency length was not relevant. Those who received instruction on the mappings outperformed learners in an implicit condition, consistent with the results of previous studies. However both groups were significantly more accurate than untrained control participants, and so the results indicated that implicit learning had occurred to a limited degree. Parallel findings were later reported by Bitan and Karni (2004) using an artificial script rather than Greek letters.

Williams (1999) exposed native English speakers to a modified version of Italian, while assessing their memory for the input sentences. A subsequent English — Italian translation task was used to detect their knowledge of five target structures: the function of the noun, the verb, and the possessive, and adjective–noun and noun–modifier agreement. Dependency length was not relevant to the first three structures, but it was a feature of both agreement patterns. Example experimental items can be seen in (8). Attention was not focused on the form in Experiment 1, creating a situation in which implicit learning would be encouraged. Nevertheless the participants acquired both agreement structures and the function of the possessive. In Experiment 2 however, input enhancements encouraged focus–on–form and an explicit–inductive or rule–search approach. Under these conditions a different pattern of results was obtained. All of the structures were acquired except for the function of the possessive and possessive–noun agreement. Note that the possessive structures were only acquired in implicit conditions, but the functions of the noun and the verb were only acquired in explicit conditions. As performance in implicit conditions was not merely a subset of that in explicit conditions, the claim that the two experiments engendered a different type of learning is supported. The input enhancements did not merely improve learning outcomes, but qualitatively altered them.

(8) a) Odio la musica modern-a
Hate-I the-F-sg music-F-sg modern-F-sg
"I hate the modern music"

b) Spesso ripariamo le sue macchine
Often repair-we the-F-pl his car-F-s
"We often repair his cars"

Saffran (2001) exposed participants to an artificial phrase structure language with the rules in (9). Some learners knew that they would be tested later on the grammatical structure. Therefore they were likely to engage explicit learning. Other participants were not given that information and were therefore more likely to be learning implicitly. After the exposure phase the participants had to distinguish grammatical from ungrammatical exemplars in a two–alternative forced choice test. The ungrammatical items violated one of the rules in (10). The first two concern the presence or absence of an item, while the remainder can be characterised as form–form contingencies, all of which appeared to be one–step dependencies. Subsequent analyses did not reveal any significant differences between the two trained groups, although both performed more accurately than the controls with three of the four one–step dependencies. Thus, the results suggested that implicit and explicit learners were able to acquire the artificial language to a similar degree. Nevertheless, as the two groups' learning outcomes were the same, the results could be interpreted as evidence that the participants uniformly engaged either in implicit or in explicit learning, regardless of the experimental instructions. Thus, definite conclusions could not be drawn.

(9)
$$S \rightarrow AP + BP + (CP)$$
$$AP \rightarrow A + (D)$$
$$BP \rightarrow \left\{ \begin{array}{lll} CP & + & F \\ E & & \end{array} \right\}$$
$$CP \rightarrow C + (G)$$

(10) a) Every sentence must contain at least one A word.
b) No sentence may contain more than one A word.
c) If there is an E word, there cannot be a CP.
d) If there is a D word, then there must be an A word.
e) If there is an F word, then there must be a CP.
f) If there is a G word, then there must be a C word.

Learners in implicit conditions are sometimes able to acquire new second language structures, although performance is often better when the input is processed for meaning than when it is memorised. Where such comparisons were available, the experiments reported so far in this section unanimously suggested that explicit learning is more effective than implicit learning. This claim has also been supported by a meta–analysis (Norris & Ortega, 2000) and a review (DeKeyser, 2003). However in most cases the amount of exposure provided in the learning phase was relatively small, and the target language was not used for natural communication. Such conditions are not ideal for implicit learning, which generally takes longer than its explicit counterpart (McClelland et al., 1995; McClelland, 1998; Norris & Ortega, 2000). One study which addressed these concerns was conducted by Morgan-Short (2007). Participants were exposed to an artificial language named Brocanto 2 with two target structures: phrase structure and noun–determiner agreement for gender — see (11). They were trained in either an implicit or an explicit condition. Those in the former were only exposed to meaningful exemplars, while those in the latter received metalinguistic explanations of the structure of the language as well as the meaningful input. The exposure was provided during a game in which the movement of playing pieces on a board was controlled using Brocanto 2. In production blocks they watched and described moves with spoken utterances; in comprehension blocks they heard a sentence and moved the pieces accordingly. The complete training session included eight hundred and eighty such trials.

(11) a)

Blom	neim-o	l-u	neep	l-i	praz
Blom-piece	square-F	the-F	neep-piece	the-M	swap

'The square blom switches the neep'

b)	Blom	neim-*e	l-u	neep	l-i	praz
	Blom-piece	square-*M	the-F	neep-piece	the-M	swap

'*The square blom switches the neep'

Grammaticality judgement tests were conducted at various stages in the procedure, during which the electroencephalogram (EEG) was recorded and event–related potentials (ERPs) calculated. The ungrammatical test items either contained phrase structure errors or incorrect noun–determiner agreement for gender. At a pre-defined lower proficiency level the explicit learners' classification responses were more accurate than those of the implicit group for agreement violations but not for the phrase structure ones. Thus, the results were consistent with previous reports of superior performance under explicit conditions. However by the end of the training period this difference had disappeared and the two groups' behavioural results were equivalent. In contrast, the event–related potentials produced in response to both types of ungrammaticality suggested an advantage for the implicit learners. The group demonstrated components typical of first language (L1) processing even at the low proficiency level. By the end of the experiment they had developed a left anterior negativity followed by a P600, the biphasal pattern that is characteristic of the processing of L1 syntactic anomalies (Luck, 2005). The explicit learners on the other hand did not evidence any statistically significant ERP components that were typical of language processing. Thus, Morgan-Short (2007) suggested that, if sufficient meaningful and communicative input is provided, implicit SLA can lead to both accurate performance and native–like processing while explicit learning only leads to the former.

Implicit learning has led to above–chance performance with a variety of structures including agreement, soft mutation and grapheme–phoneme correspondences, some of which were one–step dependencies. However, the process appears to have a smaller role in successful second language acquisition than does explicit learning, at least when exposure is limited, and success was not guaranteed. The only two–step contingency to be examined was not acquired. Overall therefore, the findings are consistent with a decline in performance coinciding with an increase in dependency length. Nevertheless, the evidence is insufficient to draw any definite conclusions because there was only a single study using a target structure with a non-adjacent contingency. Therefore the following sections report evidence from similar experiments in the artificial grammar learning, serial response time and statistical learning paradigms, in which the issues of dependency length and chunk size have been addressed more directly.

2.4 Artificial Grammar Learning

In a prototypical artificial grammar learning experiment, participants begin by viewing and often memorising letter strings. Unbeknownst to the participants these strings are generated according to a complex set of rules. The learners' resulting knowledge of these rules may be measured either as a decrease in the amount of time required to memorise new strings (e.g. A. S. Reber, 1967), or by means of a subsequent grammaticality judgement test (e.g. Altmann, Dienes, & Goode, 1995; Channon et al., 2002; Dienes & Scott, 2005).

An additional task is often used to assess awareness of the rules governing the letter strings. As in SLA this may be verbal report in the form of an interview (A. S. Reber & Allen, 1978), or of instructions given to a further "yoked" participant on how to complete the task (Mathews et al., 1989). Alternatively, explicit knowledge is sometimes assessed in generation tasks where participants are asked to produce grammatical strings (e.g. Gómez, 1997) and recognition tasks where they identify chunks that had been used in the training strings (e.g. Gómez & Schvaneveldt, 1994).

According to the evidence presented in Section 2.1, implicit knowledge is unlikely to be utilised to a large extent in verbal report and generation tasks because there is no relevant triggering input in either case. In the recognition task on the other hand the participants are presented with string chunks, making it unclear why implicit knowledge should not contribute to performance. Therefore this review does not include experiments that relied solely on recognition to differentiate implicit from explicit knowledge.

An example transitional artificial grammar used by Dienes and Scott (2005) is shown in Figure 2.1. It is used to generate strings of letters by starting at the *IN* arrow, and following the arrows until reaching *OUT*. Therefore *XMXM* would be a legal string but *XMXTM* would not be, as neither of the transitions from the second *X* passes a *T*, making the *XT* pair illegal. Thus, many of the rules governing the output of a transitional artificial grammar are one–step dependencies. Each letter determines which can legally follow. (As such they can also be described as bigrams, or two letter chunks.) However there are also longer chunks. For

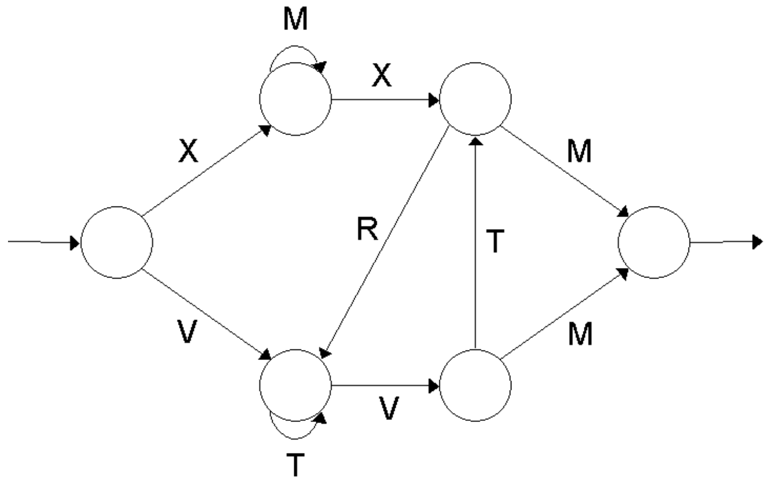


Figure 2.1: Transitional artificial grammar. Each legal letter string is represented by a path through the diagram from the *In* node to the *Out* node, always following the direction of the arrows.

example *TVV* is not grammatical, but its illegality cannot be explained in terms of the constituent bigrams as both *TV* and *VV* are generated by the grammar. Rather, it is necessary to consider the three letters as a whole. In the current terminology these are described as trigrams (three letter chunks) rather than two–step dependencies because the identity of the final letter depends not just on the first, but rather on the combination of that letter and the intervening one. This can be demonstrated by changing the intervening letter to *R*, which creates the legal sequence *TRV*.

A. S. Reber (1967) conducted the original study in the AGL paradigm. He showed that participants who were given instructions designed to engender implicit learning could memorise exemplars more quickly with practice, but only when they were exposed to rule–governed letter strings. In addition, the participants were able to differentiate grammatical and ungrammatical items in a subsequent judgement test. Thus, it is probable that they had acquired the one–step dependencies in the strings. However A. S. Reber did not investigate whether the resulting knowledge was implicit. His findings have since been replicated by other researchers including Altmann et al. (1995) and Don, Schellenberg, Reber, DiGirolamo, and Wang (2003).

A. S. Reber (1976) again investigated the contribution of implicit learning in the AGL task. However in a change to the earlier procedure, he also included a comparison group instructed to search for rules under explicit conditions. Both groups performed above chance in a subsequent grammaticality judgement test, demonstrating knowledge of the one–step dependencies governing the letter strings. Nevertheless, the implicit learners were significantly more accurate. Even if both groups engaged explicit learning to some extent, its contribution was likely to have been larger for the intentional rule–search participants than for the incidental group. Thus, if test performance was solely the result of explicit learning then the explicit group should have been the more accurate of the two. This was not the case. As such, implicit learning was likely to have contributed to the implicit group's superior performance.

It is also important to assess the extent to which performance in artificial grammar learning experiments depends on implicit knowledge. Dulany, Carlson, and Dewey (1984) first claimed that explicit knowledge identified via verbal report could fully account for the participants' judgement test performance and the results were interpreted as evidence that implicit knowledge was not employed.

Manza and Bornstein (1995) reported a judgement test in which participants either decided whether strings were grammatical, or whether they liked the items. Explicit knowledge was presumed to be utilised to a greater extent in the former task than in the latter because there was a greater focus on form in the grammaticality judgement test. Performance was significantly above chance for both tasks, and there was no difference between the two groups. Therefore, if the affective judgement group was using implicit knowledge as assumed, the experiment provided evidence of implicit knowledge. Alternatively however the manipulation may have been unsuccessful and the two groups may have used the same learning mode (that could have been either implicit or explicit). There is insufficient evidence to choose between these two interpretations.

Stronger evidence has been provided by studies of anterograde amnesiacs, who are unable to acquire new explicit knowledge. After exposure, amnesiac participants tested by Meulemans and Linden (2002) were able to classify test strings as grammatical or ungrammatical at above chance levels. However as predicted they were unable to generate legal strings, a task presumed to require explicit knowledge (Gómez, 1997). Similar results have also been reported by Don et al. (2003), studying participants with Williams Syndrome.

The studies reported so far provided evidence that learners can acquire the one–step dependencies that characterise the rule–governed strings generated by a transitional artificial grammar. Perruchet and Pacteau (1990) extended the findings to investigate whether learners can acquire implicit knowledge of larger chunks. Participants were trained either on bigrams or on complete letter strings before their resulting knowledge was tested in a grammaticality judgement test. If the group trained on the full strings outperformed those receiving only bigrams, then the former would be assumed to know more than simply the pairs of letters. However the judgement accuracy of the two groups was equivalent, with the exception that those trained on the whole strings knew which bigrams were legal in first position. Thus, the authors concluded that successful performance in AGL experiments is largely the result of bigram knowledge.

Gómez (1997) took a different approach to the same question. All participants were trained on the full strings. Unlike Perruchet and Pacteau (1990) she then excluded the participants who demonstrated explicit knowledge in a generation test, thus ensuring that only performance based on implicit knowledge was measured in the main analyses. Half of the ungrammatical judgement test items contained illegal bigrams. In the remainder the bigrams were grammatical but the trigrams were not. The participants without explicit knowledge were able to identify the judgement test strings containing ungrammatical bigrams but not those with illegal trigrams. Thus Gómez, like Perruchet and Pacteau interpreted the results as evidence that implicit knowledge is restricted to pairs of adjacent letters.

Mathews et al. (1989) tested whether participants could acquire implicit knowledge of longer–distance dependencies. They employed a different type of artificial structure: the biconditional grammar that is shown in Figure 2.2. It generates strings of eight letters, separated into two sets of four by a full stop. The first letter uniquely determines the fifth, the second the sixth, the third the seventh and the fourth the eighth as follows. If the letter in one position is *D*, then that in the linked position must be *F*. *G* is similarly linked with *L*, and *K* with *X*. Thus, a string generated by a biconditional grammar contains four overlapping four–step dependencies. Mathews et al. trained participants either on the match task in which they memorised individual items and then identified them from a list of three possibilities, or on an edit task in which they used trial and error to improve an ungrammatical string. The former task was expected to engender implicit learning and the latter was expected to encourage explicit learning because it was intentional. There was no evidence that the match–task participants had acquired the target four–step dependencies. These results have since been replicated by Johnstone and Shanks (2001) and Shanks, Johnstone, and Staggs (1997).

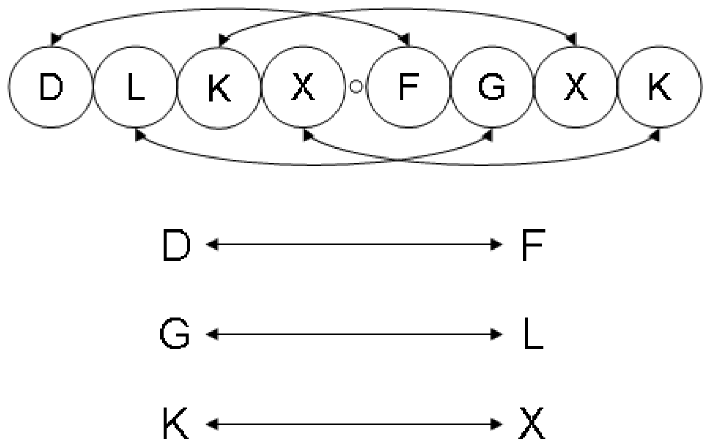


Figure 2.2: Biconditional grammar. The system produces strings of eight letters, where the first uniquely determines the fifth, the second the sixth and so on. If the first letter is *D*, the letter in the position five must be *F*. The letter pairs *G* and *L*, and *K* and *X* are similarly linked.

In summary, learners are able to acquire implicit knowledge of one–step dependencies or bigrams in artificial grammar learning experiments. However their performance is reduced with larger chunks and with longer dependencies. Trigrams and four–step dependencies do not appear to be acquired. If performance with continuous chunks is superior to that with discontinuous dependencies as claimed by Jiménez et al. (1996), it is unlikely that learners would acquire implicit knowledge of any long–distance dependencies in artificial grammar learning experiments. The following section considers whether the same applies in a different experimental task that was also designed to investigate the acquisition of implicit knowledge.

2.5 Serial Response Time Task

The Serial Response Time (hereafter SRT) task was also designed to investigate implicit learning and the acquisition of implicit knowledge (Nissen & Bullemer, 1987). Participants respond to the location of a stimulus, which can appear at one of (usually) four positions on a computer monitor, by pressing the corresponding key on the keyboard as quickly as possible. In many studies the stimulus follows a repeating sequence of six, ten or twelve locations (e.g. Howard, Mutter, & Howard, 1992; Nissen & Bullemer, 1987). In others, the location is generated according to a transitional artificial grammar of the type used in AGL research (e.g. Cleeremans

& McClelland, 1991; Jiménez et al., 1996). Knowledge of the regularities governing the stimulus location is inferred when reaction times decrease further for trained participants than for control participants who experience a random series of locations (Nissen & Bullemer, 1987). Alternatively, experiments may use within–subjects designs, contrasting performance in legal trials with that in illegal trials (Cleeremans & McClelland, 1991; Howard et al., 1992; Shanks et al., 2005). As in artificial grammar learning, a second task is used to detect explicit knowledge. Often, participants are asked to report whether they were aware of the pattern in the locations (Curran & Keele, 1993; Nissen & Bullemer, 1987; Willingham & Goedert-Eschmann, 1999) or to generate the sequence (Howard et al., 1992; Nissen & Bullemer, 1987; Willingham & Goedert-Eschmann, 1999). Alternatively the participants may be anterograde amnesiacs, who are assumed to be unable to acquire new explicit knowledge (Nissen & Bullemer, 1987; Nissen, Willingham, & Hartman, 1989; P. J. Reber & Squire, 1998).

Using these techniques, Nissen and Bullemer (1987) originally discovered that participants trained on a repeating sequence improved their reaction times more than those trained on a random series. Therefore the authors concluded that the participants had acquired the repeating sequence. However subsequent verbal reports indicated that the majority were aware of it, and the unaware participants' reaction time data were not analysed separately. Thus, explicit knowledge may have influenced the measures of SRT performance. In summary, Nissen and Bullemer demonstrated that learners had acquired the sequence, but the experiment did not provide any evidence of implicit knowledge.

More recent experiments have replicated the reaction time results while ensuring that performance stemmed at least partially from implicit knowledge. Frensch and Miner (1994) gave learners incidental or intentional instructions designed to encourage implicit or explicit learning respectively. After the SRT task, the participants completed a generation task to assess their explicit knowledge. Those trained in incidental conditions were at chance in the generation task, but examination of their SRT performance nevertheless indicated that their reaction times (RTs) were significantly faster in legal than in illegal items. Thus, the results were interpreted as evidence that the group had implicit knowledge of the regularities governing the stimulus location. Nevertheless, their SRT task performance was below that of the intentional group that demonstrated a larger contrast between RTs in the illegal and legal trials. Thus, the results were consistent with reports in the SLA literature that implicit learning and knowledge are less effective than explicit learning and knowledge (Norris & Ortega, 2000). Similar results have been reported by Willingham and Goedert-Eschmann (1999), using verbal report rather than the generation task to confirm that knowledge was not accessible to conscious awareness, and by Nissen and Bullemer (1987), Nissen et al. (1989) and P. J. Reber and Squire (1998), based on the performance of anterograde amnesiacs.

As in the artificial grammar learning paradigm, success in the original version of the SRT task could result from knowledge of very simple stimuli characteristics (Curran, 1997; Reed & Johnson, 1994a; Shanks & St. John, 1994). The original ten–location sequence used by Nissen and Bullemer (1987) was *4231324321* (where *1* represents the leftmost of the four locations and *4* the rightmost). In each single repetition of the sequence positions *1* and *4* are encountered twice, and positions *2* and *3* three times. Thus, participants could perform at above chance levels simply by knowing that at any point the latter pair of responses was more probable than the former.

Knowledge of the adjacent pairs of locations would also have led to faster RTs on legal than on illegal trials. For example, in the same sequence, *1* is never followed by *2*, *3* is never followed by *4* and *4* is never followed by *1*. Participants may therefore learn that *3* and *4* are more likely to follow *1* than is *2*. Overall, even though participants are able to respond more quickly to consistent than inconsistent trials in the SRT task, this may be the result of knowledge of low–level regularities in the data rather than of the entire sequence. Following this realisation, a further set of studies investigated whether implicit knowledge of larger chunks and longer contingencies can be acquired in SRT experiments.

Cohen, Ivry, and Keele (1990) trained participants on one of three types of target structure. Simple sequences could be described only using bigrams, as in the example in (12*a*). In the example, *1* is always followed by *4*, *4* by *5*, *5* by *3*, *3* by *2*, and *2* by *1*. Hybrid sequences included both bigrams and trigrams, as in (12*b*). In the example, *4* was always followed by *2* and *3* was by *1*. However, *1* could be followed either by *4* or *2* depending on the context, while *2* could be followed by either *3* or *1*. On the other hand the ambiguous sequences could only be described using trigrams, as in (12*c*). In these it was always necessary to consider the two preceding locations in order to uniquely determine the subsequent one. In all cases the participants' generation task performance suggested that they were not aware of the regularities. This pattern of results indicated that implicit knowledge of both bigrams and trigrams can be acquired in the SRT task.

(12) a) 14532
b) 142312
c) 132312

The claim that trigrams can be acquired has since been supported by Curran and Keele (1993), Frensch, Buchner, and Lin (1994), Jiménez et al. (1996), Rowland and Shanks (2006) and Shanks et al. (2005). However there is less consensus about whether implicit or explicit knowledge is involved. The findings of Curran and Keele and Frensch et al. remained unchanged after participants with above chance verbal report or generation task performance were removed from the analyses, while Jiménez et al. reported that knowledge of some trigrams was only apparent in the SRT task and not in generation. Thus, the three studies produced evidence of implicit knowledge. In contrast, Shanks et al. claimed that generation task performance (and therefore explicit knowledge) was sufficient to account for their participants' improvements in the SRT task. Curran (1997) also reported that anterograde amnesiacs (who can only acquire new implicit knowledge) were unable to acquire sequences depending on trigrams. Overall therefore, implicit knowledge of trigrams may be acquired, but it does not occur reliably in all experiments.

Cleeremans and McClelland (1991) carried out a detailed investigation of the influence of chunk size on the acquisition of implicit knowledge in the SRT task. The stimulus locations in Experiment 1 were specified according to an artificial grammar such that the input contained relevant two–, three–, four– and five–item chunks, knowledge of which could speed performance on the SRT task. The participants gradually became sensitive to the two– three– and four–item chunks, but not to the five–item chunks even after sixty thousand exposure trials. In a second experiment they tested the acquisition of four–step dependencies, where the intervening elements were not of predictive value. Despite undergoing the same lengthy training as in Experiment 1, their SRT performance provided no evidence that the participants had acquired this target structure.

In summary, implicit knowledge of four–step dependencies is not acquired in the SRT task as in artificial grammar learning. However, shorter long–distance dependencies were not investigated in either case, so no precise limits can be specified. Turning to continuous chunks, unimpaired participants appear to acquire implicit knowledge of bigrams and trigrams while carrying out the SRT task, although performance with trigrams may be reduced in amnesia. In contrast, there is no evidence that five–item chunks are acquired even when substantial input is provided. These findings are consistent with the claim that implicit knowledge of longer structures cannot be acquired, as was also proposed for artificial grammar learning in Section 2.4. However, the chunk size limit is different in each paradigm. Implicit knowledge of trigrams was not acquired during exposure to the output of an artificial grammar, but it was during the SRT task. There are too many differences between the two tasks to pinpoint the cause of this contrast. Nevertheless, it is important to note that chunk length, like dependency length, affected performance in both paradigms.

2.6 Statistical Learning

Research in the statistical learning paradigm began with a study by Saffran, Aslin, and Newport (1996). They exposed eight–month old infants to a two–minute–long continuous stream of artificially generated nonsense words, all of which were trisyllabic. As the synthesised speech did not include acoustic cues to word boundaries, the only way to segment the stream was to use the transitional probabilities between adjacent syllables. On every presentation, the first syllable of a word was obviously followed by the second, which itself was always followed by the third. However, as the third syllable could be followed by the start of any of the other three words, it was only succeeded by any particular one 33% of the time. Thus, at word boundaries there was a reduction in the probability of the transition between syllables. As this structure concerned relationships between adjacent syllables, it can be characterised as a one–step dependency.

After exposure to these speech streams, the infants heard a repeated series of three syllables taken from the exposure stream. Some of these items corresponded to the nonsense words, while others (termed part–words) consisted of the final syllable of one word and the first two syllables of another. The researchers then calculated the amount of time that the infants oriented towards the loudspeaker that was playing the test stimulus. The results demonstrated that the participants looked significantly longer in part–word trials than in word trials, and therefore that they were able to distinguish the two types of stimuli. These findings were interpreted as evidence that the infants had acquired sensitivity to the transitional probabilities between adjacent syllables. The remainder of the section discusses whether dependency length affects performance in statistical learning experiments as it does in the other experimental domains discussed above. It primarily focuses on investigations of statistical learning in adults rather than in infants.

Unlike the other paradigms summarised above, experiments in statistical learning do not tend to address the issue of awareness or differentiate between implicit and explicit knowledge. For infants the task is clearly incidental and therefore likely to engender implicit learning. Adult participants on the other hand are sometimes instructed to identify the word boundaries in the input, which would encourage an intentional or explicit learning mode (e.g. Saffran, Newport, & Aslin, 1996). However Saffran, Newport, Aslin, Tunick, and Barrueco (1997) reported consistent results when focusing participants' attention on a cover task, drawing pictures while the speech stream played in the background. The similarities between adult and infant performance also indicate that implicit learning may contribute to adult performance. Thus, although awareness is not usually assessed in statistical learning experiments, research in the paradigm is recognised to be relevant to implicit learning and the acquisition of implicit knowledge (Cleeremans et al., 1998; Gómez & Gerken, 1999).

Saffran, Newport, and Aslin (1996) demonstrated that adults were able to acquire one–step dependencies in a similar environment to the infants reported in Saffran, Aslin, and Newport (1996). The participants were asked to identify the word boundaries during their twenty–one minutes of exposure to a continuous speech stream. Their resulting knowledge was assessed in a two–alternative forced–choice test in which the participants were instructed to distinguish the nonsense words from part–word foils. They performed at above chance levels, and therefore were also able to segment the speech stream based on one–step dependencies between syllables. This finding has since been replicated with infant, child and adult participants (Aslin, Saffran, & Newport, 1998; Gómez & Gerken, 1999; Saffran et al., 1997; Thiessen & Saffran, 2003; Toro, Sinnett, & Soto-Faraco, 2005) and similar performance has been reported with sequences of tones rather than of syllables (Saffran, Johnson, Aslin, & Newport, 1999). Overall therefore, the ability to acquire one–step dependencies in the statistical learning paradigm appears to be robust (see Gómez & Gerken, 2000; Saffran, 2003, for reviews).

Gómez (2002) asked whether variability in the intervening element affected the acquisition of two–step dependencies. The experimental materials were designed so that the first element in each item could be followed by any second element, but that the identity of the third was uniquely determined by the first. The set of second elements contained two, six, twelve or twenty–four items. The input was segmented, with 250 ms gaps between elements and 750 ms ones following each item. After eighteen minutes of exposure to the speech stream, the adult participants had acquired sensitivity to the target two–step dependency but only in the twenty–four set condition. Thus, Gomez reported that learners are able to acquire two–step dependencies when there is a large degree of variation in the identity of the intervening element.

Peña et al. (2002) also investigated whether adults could segment the speech stream based on two–step dependencies. Participants were trained on a continuous stream of trisyllabic nonsense words for ten minutes. In each word the first syllable uniquely determined the third, but the second varied freely between three options. A subsequent test demonstrated that the learners could distinguish nonsense words from part–words, indicating that they had acquired the target two–step dependency. As Peña et al. only used a set of three second syllables, these findings were apparently inconsistent with those of Gómez (2002). Nevertheless, it is possible to reconcile the results of the two studies. Gómez used disyllabic intervening elements. Despite the cues to segmentation provided by the pauses in the stream, the participants may have processed the input as a sequence of syllables and therefore perceived the target contingency as containing three steps. On the other hand, Peña et al. only employed monosyllabic second elements, and therefore definitely used a two–step dependency.

In a more recent study, Newport and Aslin (2004) tested whether longer exposure periods would improve adults' performance with two–step dependencies presented in a continuous stream. Within each trisyllabic word, the first syllable was selected from a set of five, the second was freely chosen from a different set of four, and the third was uniquely determined by the first. Thus, there was a two–step dependency between the first and third syllables, while the second did not provide as much predictive information. Under these circumstances the participants were not able to distinguish words from part–words following twenty–one minutes' exposure, or even after ten such training sessions spread over consecutive days. Thus, the authors concluded that statistical learning processes cannot be used to acquire two–step dependencies between non-adjacent syllables in continuous speech[3]. The results were consistent with Gómez (2002), as Newport and Aslin only employed four intervening elements. However the

[3]Note that Newport and Aslin (2004) also conducted further experiments in which the participants acquired a two–step dependency between non-adjacent phonemes. However, if they segmented the input into syllables rather than phonemes, the target patterns would actually have been one–step dependencies. As there were no pauses in the speech stream, the participants were not given any guidance as to how to segment the input so this possibility cannot be ruled out. Thus, these experiments do not provide any firm evidence that non-adjacent contingencies can be acquired in statistical learning experiments.

findings stand in contrast to Peña et al. (2002). One possible explanation is that the syllables used by Peña et al. were more distinctive than the ones used by the other researchers, as they all contained different initial consonants. This may have simplified the learning task, making participants less likely to confuse elements.

In summary, research in the statistical learning paradigm has focused on long–distance dependencies rather than on continuous chunks. There are robust findings that participants in statistical learning experiments can acquire one–step dependencies between adjacent syllables. Performance can be maintained with two–step dependencies but only under limited circumstances partially determined by the distinctiveness of the individual elements, and the length and variability of the intervening element. More research is required to clarify the necessary conditions. Nevertheless, although two–step contingencies are sometimes acquired, the reduction in performance relative to one–step versions is consistent with the interpretation that success in statistical learning experiments is restricted to shorter structures.

2.7 Other Tasks

Although many researchers have focused on the AGL and SRT tasks or on statistical learning, some important experiments were based on different techniques. For example, citetPacton2008 used a learning task in which participants carried out calculations based on two digits in a sequence, thus forcing them to pay attention to those digits. Unbeknownst to the learners there was also a single one– and a single two–step dependency in the digit sequence. A subsequent two-alternative forced choice test was used to ascertain whether the participants had acquired either of those contingencies, although there was no measure of whether the resulting knowledge was implicit or explicit.

The results suggested that the learners could acquire either the one– or the two–step dependency but only when they had to focus on the relevant digits as part of the calculation task. If they were not required to pay attention to those particular digits they did not acquire either contingency. Thus, the authors concluded that simultaneous attention to both elements rather than their adjacency is the crucial factor in determining whether a dependency between them can be acquired.

This finding is actually consistent with the results reported earlier in the chapter. Recall that the argument given to support the role of adjacency in Section 2.2 was also based on attention. If the task does not force the participants to focus on elements in a different order then the learners would pay attention to them in the order in which they were presented, and an adjacency effect would be detected. The tasks normally employed in the AGL, SRT and statistical learning paradigms generally do not involve such manipulations of attention. When speakers process a known language they may notice the first element in a potential dependency and hold it in working memory until the second element is encountered. However learners beginning to acquire a novel language would be less likely to notice these structures, instead processing each element of the input in the order in which it was presented.

2.8 Research Questions

By definition, implicit and explicit knowledge differ in terms of awareness (see Section 2.1). However there is debate about how to operationalise this in order to detect one form to the exclusion of the other. A variety of measures were employed in the experiments reported in Sections 2.3, 2.4 and 2.5. Production tasks and timed grammaticality judgement tests were usually used to assess implicit knowledge, while untimed grammaticality judgements, generation tasks and verbal report tended to be utilised to detect explicit knowledge. However, claims that a test is biased towards either form of knowledge are controversial, and assertions that performance on a task is purely the result of either type are even more so (Gaillard, Vandenberghe, Destrebecqz, & Cleeremans, 2006; Shanks & St. John, 1994). Therefore the first research question is whether we can develop an improved method of distinguishing implicit from explicit knowledge.

Although significant implicit knowledge has been detected in some studies of second language acquisition (R. Ellis, 2005; Erlam, 2006; Hulstijn & Hulstijn, 1984; Seliger, 1979; Williams, 2005), it has not in others (Han & Ellis, 1998; E. Hauser, 2000). Thus, accurate implicit knowledge of linguistic target structures does not develop reliably (DeKeyser, 2003). In addition, when it was detected it was often by means of verbal report, a procedure that may overestimate the role of implicit knowledge (Shanks & St. John, 1994). It is important to replicate the findings while using a better method to differentiate implicit from explicit knowledge, in order to confirm that implicit knowledge can be acquired. This is the focus of the second research question.

The third research question is the main focus of this thesis. It asks whether the acquisition of implicit knowledge of a second language structure is affected by the length of the target dependency when the learning task does not manipulate the order in which elements are held in attention. Specifically, it considers whether an increase in the number of words intervening between a noun and a verb reduces the likelihood of acquiring an agreement relation between the two.

In summary therefore, this thesis addresses the three research questions listed in (13).

(13) a) Is it possible to exclude explicit knowledge in order to obtain a valid measure of implicit knowledge?
b) Is there evidence of the acquisition of implicit knowledge in second language acquisition?
c) Does dependency length limit the acquisition of implicit knowledge of a second language?

2.9 Chapter Summary

This chapter first defined implicit knowledge and identified some of its key characteristics, including the assumption that it develops automatically and continuously. Next, previous research into implicit knowledge and implicit learning was summarised with the tentative conclusion that both contribute to second language acquisition and use to a limited extent. Thus, there is a contrast between theoretical claims that implicit knowledge is acquired continuously and empirical findings that second language structures are not always acquired.

One explanation that could account for the contrast is that the acquisition of implicit knowledge is restricted to relatively simple target structures and specifically to shorter dependencies. Supporting this claim, evidence was presented from other experimental paradigms (artificial grammar learning, the serial response time task and statistical learning) in which contingency length limits performance. The proposed explanation was also shown to be consistent with previous results in the second language acquisition literature, although as there was very little research looking at longer structures firm conclusions could not be drawn. Finally, we presented a series of research questions that directly addressed the effects of dependency length on the acquisition of implicit knowledge in adult SLA.

CHAPTER 3

Experimental Design

Five experiments will be reported in this thesis, all of which used the same basic method but with variation in the specific learning tasks and target structures. This chapter first provides an overview of the common aspects of the experimental design. Then it discusses three important issues that had to be addressed before the experiments were planned. Note that detailed methodological information specific to each experiment can be found in the method sections of the following chapters (Sections 4.1, 5.1, 6.1, 7.1 and 8.1).

3.1 Overview

Each experiment consists of a learning phase followed by a test phase. In the former, participants are exposed to either the output of an artificial biconditional grammar or to sentences in a novel foreign language. During that exposure phase participants complete one of two memory tasks (reproduction and source localisation) described in Sections 4.1.3 and 6.1.3 respectively. For quick reference, Table 3.1 lists the tasks and structures that were used in each experiment.

In the biconditional grammar experiment (Experiment 1, see Chapter 4) the target structures are a series of four–step dependencies between letters in a string. (Remember that a four–step dependency is a contingency between two elements in a sequence that are separated by three intervening elements.) In the linguistic experiments the target structure is either subject–verb or object–verb agreement for number (with the former allowing transfer from the participants' L1 and the latter not). The number of additional words intervening between the relevant noun and the finite verb is manipulated between experiments. (See the following section for more discussion of how this was measured.) The learning task was not designed to influence the order in which the separate letters or words were held in attention.

There are three sections to the test phase in each of the experiments. The first is a timed grammaticality judgement test (hereafter the timed test) and the second is an untimed grammaticality judgement test (hereafter the untimed test). Two variables (grammaticality and familiarity) are manipulated in each of these tests as follows. In the ungrammatical stimuli there is an error in the target structure. Familiar items only contain bigrams (adjacent word pairs) that are seen in the learning stimuli while unfamiliar items contain at least one novel bigram. In other studies familiarity has often been manipulated in addition to grammaticality, to ensure that if participants respond on the basis of an item's constituent bigrams, their performance cannot masquerade as an effect of grammaticality (e.g. Channon et al., 2002; Johnstone & Shanks, 2001).

The final element is a multiple–choice correction test based on the rule–choice test used by R. Ellis (2005). Incorrect items are presented alongside four suggested corrections. Participants are instructed to select the option that would improve the item, making it grammatical. The results of this task are used to divide the trained participants into three groups: rule detectors (who are defined as responding to a sufficient number of items by choosing the option that would make them grammatical), familiarity detectors (who select the option that increases the number of frequent bigrams in a sufficient number of items) and non-detectors (who do not reach the criterion level of performance). No familiarity detectors can be identified by the correction test in Experiments 2 and 3, because none of the four options increase the number of familiar bigrams in the incorrect item.

The judgement tests are analysed separately for each group of trained participants. Their performance in these tests is compared with that of untrained control participants, to determine whether the former acquire either the relevant target structure or the familiar bigrams. The focus is on the non-detectors, but the results of the other two groups are also analysed for comparison.

3.2 Measuring Dependency Lengths

The evidence presented in Sections 2.4 and 2.5 suggested that participants in artificial grammar learning and serial response time experiments can acquire

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
Materials	Biconditional	Persian	Persian	Basque	Basque
Structure	N/A	S–V	S–V	O–V	S–V
Length	4 Steps	2 Steps	2 Steps	2 Steps	3 Steps
Number of Dependencies	4	1	1	1	1
Task	Reprod.	Reprod.	Source	Source	Source

Table 3.1: Design of experiments. The table shows the materials, target structures and learning tasks employed in each of the five experiments in the thesis. S–V = subject–verb agreement; O–V = object–verb agreement; Reprod. = reproduction task; Source = source localisation task.

bigrams / one–step dependencies and sometimes also trigrams, but not longer contingencies. Adults and infants also appear to be able to use one–step dependencies when segmenting continuous speech streams in the statistical learning paradigm, but performance is often reduced with two–step dependencies (see Section 2.6). However whether a structure is classified as a one–, two– or more–step contingency depends on the definition of an element, and this varies between research fields (see Newport, Hauser, Spaepen, & Aslin, 2004, for a different discussion of these issues).

In artificial grammar learning experiments participants are exposed to rule–governed letter strings. The length of each dependency is measured according to the number of intervening letters. In the statistical learning paradigm participants are exposed to synthesised speech streams which do not contain any relevant prosodic information. Sometimes these streams are described as a series of multisyllabic words and therefore the length of a dependency is expressed according to the number of intervening syllables (e.g. Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996). Others consider the streams to be a series of sentences containing (usually) monosyllabic words (e.g. Gómez & Gerken, 1999; Marcus, 1999a), and the length of each contingency is therefore dependent on the number of intervening words. In the absence of prosodic information a series of three syllables could be either a trisyllabic word or a three–word utterance where each word happens to be monosyllabic. Therefore this uncertainty in nomenclature rarely affects the calculation of contingency length in statistical learning experiments. However, it means that these studies do not provide any evidence as to whether participants segment linguistic input into word– or syllable–level units. This distinction is crucial for the current experiments, because the materials use multisyllabic words.

The stimuli in the current experiments were pre-segmented into words. Naturally there were spaces between words in the written versions, and there were also short pauses in the spoken items (see Sections 5.1.2 and 7.1.2). Thus, the participants were encouraged to segment the input into word–level units (Shanks & Johnstone, 2002). For this reason we adopt the approach taken by Robinson (2005) and report dependency lengths based on the number of intervening words.

3.3 Distinguishing Implicit and Explicit Knowledge

This thesis is focused on implicit knowledge. In particular it asks whether adult participants can acquire implicit knowledge of a second language, and whether their ability to do so is constrained by dependency length. To answer these questions one of the major hurdles was to identify implicit knowledge, and therefore to distinguish it from explicit knowledge. This section outlines the approach taken.

Implicit and explicit knowledge are usually distinguished in the test phase. Often one test is designated a test of implicit knowledge and another of explicit knowledge. However, as discussed above, there is disagreement about the extent to which any one test can detect knowledge of one type without intrusion of the other. Therefore, the test phase in the current experiments was based on R. Ellis (2005, 2006) and Han and Ellis (1998)[1]. R. Ellis (2005) examined the performance of existing learners of English as a foreign language on a variety of English structures. They were tested using an oral-narrative testan imitation test, a timed grammaticality judgement test, an untimed grammaticality judgement test, and a test of metalinguistic knowledge. In the first task the participants listened to and repeated incorrect sentences. If they spontaneously repaired an item, they were understood to know the rule that had been broken. There were two parts to the metalinguistic test (both multiple choice). In the first, participants selected the rule that had been violated in an ungrammatical sentence; in the second they identified specific grammatical features (such as a finite verb or a preposition) in an English text.

R. Ellis then carried out a factor analysis to compare the participants' performance on the five tests. The analysis demonstrated that the oral narrative and imitation

[1] Isemonger (2007) highlighted some statistical concerns with the original analyses employed by R. Ellis (2005), including the fact that he used an exploratory factor analysis instead of the more appropriate confirmatory version. R. Ellis and Loewen (2007) reanalysed the original data to address these issues, reporting results consistent with the original study.

tasks and the timed grammaticality judgement test loaded onto one factor, while performance on the untimed judgement test (particularly with the ungrammatical items) and on the metalinguistic test loaded more heavily onto a second. Therefore he concluded that the two groups of tests measured different types of knowledge, with the former fitting the predicted characteristics of implicit knowledge and the latter of explicit. Similar results had been reported earlier by Han and Ellis (1998), testing only a single target structure.

R. Ellis's work highlighted three tests that mainly appeared to detect performance resulting from implicit knowledge and two that primarily measured performance based on explicit knowledge. The distinction between the timed and untimed grammaticality judgement tests was particularly useful because the discrimination was the same in both cases (Jiménez et al., 1996). That is, in both cases participants were asked to decide whether or not a given item was correct. Therefore the timed and untimed judgement tests from R. Ellis (2005) were adopted for the current experiment.

Other researchers have made theoretical claims that explicit knowledge is relatively difficult to access under time pressure, whereas automatic access to implicit knowledge is not affected in the same way (Berry & Dienes, 1993; DeKeyser, 1995). Independent experimental evidence also supports the claim. Learners who had received metalinguistic instruction have been shown to take longer to respond in post–tests than those without such instruction (Bitan & Karni, 2004; Morgan-Short, 2007). Time constraints can also hinder access to explicit but not implicit knowledge (Bialystok, 1979). The time limit used in the current series of experiments was 2000 ms. It was stricter than that employed in Bialystok (1979) and Han and Ellis (1998), and it was within the range used by R. Ellis (2005). Thus, access to explicit knowledge was likely to be restricted, although it may not have been completely eliminated. In contrast, the untimed version was assumed to measure both performance based on explicit knowledge and that resulting from implicit knowledge. Nevertheless, the effects of the former are likely to be larger for participants with relevant explicit knowledge (Cleeremans & Jiménez, 2002).

The final element in the test phase was the correction test. It was based on the multiple–choice rule test, which was the first part of the metalinguistic task used by R. Ellis (2005). The original version was shown to load on to the same factor as the untimed test in the factor analysis, and so it was assumed to measure primarily performance based on explicit knowledge. Participants in the original version of the task were shown an ungrammatical sentence, and they selected which rule had been violated. Remember that they were experienced formal language learners, who were therefore familiar with the metalinguistic terminology used in the test. The current experiments employed linguistically naive participants, who would not have had the vocabulary to complete the same task. Therefore the test was modified so that the incorrect items simply had to be improved. Rather than identifying the rule which had been broken, participants had to select the appropriate correction from a set of four.

As well as the results of R. Ellis (2005) there are additional reasons to claim that performance in the correction test is mainly the result of explicit knowledge. Firstly, the test was not timed (Han & Ellis, 1998). Secondly, based on the assumption of automatic access, implicit knowledge should be used whenever the appropriate triggering input is present. Equally however it should not be accessed in the absence of the relevant trigger (Cleeremans & Jiménez, 2002; Squire et al., 1994). Participants were asked to select one of four corrections to a series of ungrammatical items in the correction test (e.g. *The last word should be x*). However they were never shown the output of the suggested corrections. Effectively therefore the participants judged the grammaticality of four output sentences that were not presented on the screen. As a result, any relevant implicit knowledge may not have been triggered. Finally, Scheffler and Cinciala (2008, September) asked language learners to correct ungrammatical L2 sentences, and to produce a rule that could account for the error in the original experimental item. In 80% of the cases when the correction was accurate, the rule was also correctly produced. Thus, correcting items is assumed to be strongly correlated with the presence of explicit metalinguistic knowledge.

The three tests were always taken in the same order in the current study, with the test expected to measure performance based on implicit knowledge before those focused on explicit knowledge. This minimised the amount of incorrect input that the participants had received prior to the test of implicit knowledge. Thus, both the accuracy of their implicit knowledge and the likelihood of detecting it were maximised.

As stated above, the timed test was biased towards measuring performance resulting from implicit knowledge, and the other tests were biased towards explicit knowledge. Nevertheless, no task can be assumed to measure one type exclusively (Jacoby, 1991; Shanks & St. John, 1994). For example, although access to explicit knowledge may be restricted under time pressure, it may occur to some degree. Rather than relying on a single criterion to identify performance that could be the result of implicit knowledge, three different criteria were employed in the current experiments. Only when all were met were participants classified as having relevant implicit knowledge. When none were met performance was classified as the result of explicit knowledge. No firm conclusions could be drawn when one or two of the criteria were met.

The first criterion was used to exclude participants with relevant explicit knowledge. If these learners had been included in the analyses, their explicit knowledge would have competed with and possibly masked their implicit knowledge (Cleeremans & Jiménez, 2002). Excluding participants who demonstrated relevant explicit knowledge therefore reduced the possible intrusion of explicit knowledge into the measures of implicit knowledge.

As discussed above, performance on the correction test was assumed to be based largely on explicit knowledge. Therefore, in order to exclude participants with explicit knowledge, those who performed at above chance levels (the rule and familiarity detectors) were separated from those who did not (the non-detectors). A similar approach was taken by E. Hauser (2000), Hulstijn and Hulstijn (1984) and Williams (2005). Thus, the first criterion for classifying participants as having implicit knowledge is that they performed at chance in the correction test (they were non-detectors). Note that the use of this criterion does not imply that a single participant could not have both implicit and explicit knowledge of the same structure, nor does it make any assumptions about how the two types of knowledge could interact. Rather, the criterion was used simply because the presence of relevant explicit knowledge makes it more difficult to isolate performance based solely or even mainly on implicit knowledge.

The next criterion related to the time constraints that applied in the first judgement test. As described above, explicit knowledge is believed to take longer to access, so accurate performance under time constraints is more likely to be the result of implicit knowledge. The second criterion therefore is that performance was above that of untrained controls in the timed judgement test. (Note that participants may also be able to access highly automatised explicit knowledge under time constraints. However as performance based on automatised knowledge would not meet the other criteria, it would not be mistakenly classified as the result of implicit knowledge.)

The final criterion emerged from the assumption that implicit knowledge develops automatically (see Section 2.1 above). If such learning proceeds automatically

Paper	Exp	Population	Items	Presentations
Channon et al. (2002)	Exp 1	Adult	36	144
Don et al. (2003)	Exp 1	Adult	16	48
Gómez (1997)	Exp 2	Adult	18	54
Johnstone and Shanks (2001)	Exp 3	Adult	24	144
Lieberman et al. (2004)	N/A	Adult	23	46
Poletiek (2002)	Exp 1	Adult	56	56

Table 3.2: Exposure phase lengths. The table shows the number of unique learning items and the total number of presentations in the learning phase of AGL experiments. (In some tasks the learning items were presented more than once within a single trial, so the number of presentations does not always equal the number of trials.) Successful learning occurred in all of the studies selected, although the authors did not always specify whether the resulting knowledge was implicit or explicit.

and continuously then it must also persist during the judgement tests, despite the participants knowing that much of the input at that point would be incorrect. The effect of this incorrect input would depend on the ratio of incorrect to accurate input provided. Should sufficient incorrect input be provided, the participants' representations of the regularities in the data would alter to include the new forms. As a result, their knowledge would become less representative of the target structure as a judgement test progresses, and their performance would become less accurate. Along similar lines, Don et al. (2003) found that performance based on implicit knowledge deteriorated during a judgement test, while Dienes, Broadbent, and Berry (1991) and Gómez and Lakusta (2004) reported that incorrect input in the learning phase reduced performance in an artificial grammar learning and a statistical learning experiment respectively. Therefore the third criterion for performance to be classified as the result of implicit knowledge was that it was only more accurate than that of untrained controls in the first block of a judgement test and not in the second block.

3.4 Length of Training

Relatively short exposure periods are often used in artificial grammar learning experiments in comparison with the SRT paradigm. A selection of studies is shown in Table 3.2. In this sample, the mean number of unique items in the artificial grammar experiments was twenty–nine, with a mean of eighty–two presentations of learning stimuli (including repetitions). In contrast, Cleeremans and McClelland (1991) employed sixty thousand learning trials in SRT experiments.

Exposure periods in implicit second language acquisition research tend to be closer to those in artificial grammar learning than those in the SRT task. However, participants trained in implicit conditions are often not compared either against chance or against untrained controls, so it is hard to ascertain whether learning was successful in those circumstances. Nevertheless, one example suggested that linguistic target structures can be acquired following relatively small amounts of exposure. Robinson (1997b) used forty items with no repetitions when training participants on two English rules. The participants trained in the implicit condition responded to approximately 65% of items correctly in the subsequent test. This figure is above the 50% chance level, but the significance of the effect was not tested statistically and nor was a measure of dispersion reported. Thus, although the results suggest that learning occurred in implicit conditions despite a brief exposure period, firm conclusions cannot be drawn.

There are advantages to a relatively short exposure period. Practically, it can be easier to recruit and retain participants, without the dropout that can occur when an experiment is spread over multiple sessions. More importantly however, it maximises the distinction between performance based on implicit and explicit knowledge. As noted above in Section 3.3, one of the crucial differences is that the latter takes longer to access. However large amounts of practice increase the speed with which explicit knowledge can be used (DeKeyser, 1997). Equally, the greater the amount of correct input in the learning phase, the less susceptible any resulting implicit knowledge would be to change following exposure to the relatively small amount of incorrect input provided in the judgement tests. Large amounts of exposure would therefore neutralise two of the three contrasts between implicit and explicit knowledge that were used to create the experimental criteria. For these reasons we chose to use an intermediate learning phase closer to the lengths usually employed in artificial grammar learning than to serial response time experiments. The number of learning items was above the median value for the experiments cited in Table 3.2, while the number of presentations was higher. Specifically, there were twenty–four unique items in each of Experiments 1, 4 and 5, and forty in Experiments 2 and 3. Each item was repeated eight times, making a total of one hundred and ninety–two trials in Experiments 1, 4 and 5, and three–hundred and twenty in Experiments 2 and 3.

Implicit knowledge develops relatively slowly, so even an intermediate exposure period may create difficulties (McClelland, 1998). This may be a problem in the current project because it aims not only to identify accurate implicit knowledge but also to investigate the circumstances under which it can and cannot develop. Answering the latter research question relies on negative as well as on positive findings. If the participants do not acquire implicit knowledge of one of the target structures, it could either mean that the learning period was too short, or else that the structure itself was not amenable to implicit learning. Note however that this problem remains regardless of how much exposure is provided.

There is some evidence from longer experiments that performance may not improve when the exposure is substantially increased. In a statistical learning experiment, Newport and Aslin (2004) found that adult participants were unable to acquire a two–step dependency between syllables in a continuous speech stream. This was counter to the original prediction, and therefore the authors tested the finding in subsequent experiments by increasing the exposure period ten–fold from one thousand five hundred and twelve trials to fifteen thousand one hundred and twenty. Nevertheless, there was no evidence that the target structure had been acquired. Consistent evidence was provided by Cleeremans and McClelland (1991), who reported SRT experiments using sixty–thousand exposure trials, in which bigrams and trigrams but not larger chunks were acquired. These findings were consistent with results from experiments with shorter learning periods. Frensch et al. (1994), Reed and Johnson (1994b) and Schvaneveldt and Gomez (1998) all provided evidence that participants can acquire two–step dependencies during the SRT task, despite using approximately one thousand five hundred exposure trials.

The effect of lengthening the exposure phase was estimated in this thesis by using connectionist simulations to model the results of each experiment (see Chapter 10). Such simulations have frequently been used to replicate and extend experimental findings in the implicit learning literature (see N. C. Ellis, 2003, for a review). As a connectionist network does not have a mechanism for awareness, it is possible to increase the exposure period infinitely without the risk of confusing implicit and explicit knowledge. In cases where the original experiment did not produce any evidence of implicit knowledge, an extended simulation tested whether it could develop from the input statistics following sixty–thousand learning trials. This is the same number as was employed by Cleeremans and McClelland (1991). Nevertheless, although they provide intriguing hypotheses for testing in future experiments, such models should not be used to draw firm conclusions about the abilities of human learners.

3.5 Chapter Summary

Participants in five experiments were trained on either the output of a biconditional grammar or on foreign language sentences. The target structure in the language experiments was noun–verb agreement. A subsequent test phase was then used to assess whether they had acquired implicit knowledge of the regularities in the input, according to three criteria.

Chapter 4

Experiment 1

A series of studies has shown that participants trained on the output of a biconditional grammar do not acquire implicit knowledge of the target four–step dependencies (Johnstone & Shanks, 2001; Mathews et al., 1989; Shanks et al., 1997), see Section 2.4 for a more detailed discussion. In contrast, those exposed to transitional artificial grammars do show evidence of implicit knowledge (Gómez, 1997; A. S. Reber, 1967, 1989; Robinson, 2005). One difference between the two types of grammar is the length of target contingency: most regularities in the output of transitional artificial grammars apply to adjacent letters, whereas biconditional grammars produce only four–step dependencies. Thus, the results suggest that learners may not be able to acquire long–distance dependencies implicitly.

Biconditional and transitional grammar experiments usually also employ different learning tasks. Participants are often trained on transitional grammars using the reproduction task, (A. S. Reber, 1976; A. S. Reber & Lewis, 1977) or by observing the input strings (A. S. Reber & Allen, 1978; A. S. Reber, Kassin, Lewis, & Cantor, 1980). However, those acquiring biconditional grammars are usually trained using a match task (Johnstone & Shanks, 2001; Mathews et al., 1989). In this procedure, participants view a string, and then identify it from a list of three, with the foils being ungrammatical. Successful learning of a transitional grammar using the match task was reported in Shanks et al. (1997) but the reverse, using the reproduction task with a biconditional grammar, has not been previously tested. That was the aim of the first experiment.

This experiment was a conceptual replication of Experiment 2 from Johnstone and Shanks (2001). The original version failed to demonstrate successful implicit learning of the long–distance dependencies generated by a biconditional grammar. Instead, the participants became sensitive to the frequent bigrams, an effect which Johnstone and Shanks claimed to be explicit according the results of a separate experiment. The current experiment used the same biconditional grammar and stimuli. However during the learning phase, participants carried out the reproduction task rather than the match task. If the structure itself is problematic for learners, then the participants should still not acquire implicit knowledge of the target rules. However if only the learning task prevented successful implicit learning in the experiments reported by Johnstone and Shanks, then implicit knowledge of the target structure should develop in the current experiment. Thus, Experiment 1 was conducted in order to confirm whether the length of a contingency affects the acquisition of representative implicit knowledge in artificial grammar learning experiments.

4.1 Method

4.1.1 Participants

Two groups of sixteen participants took part in the experiment (one trained group and one untrained control group). All were students and recent graduates of the University of Edinburgh recruited from advertisements placed in the university's Student Employment Service. To ensure that they were sufficiently motivated, a prize of a bottle of wine (or the financial equivalent) was offered for the best performance within each group.

The participants were monolingual native English speakers with minimal exposure to any other language before age eleven. A questionnaire was used to collect further information on the number of languages they had studied after that age, and their current ability in their best foreign language, self–reported on the seven–point scale shown in (14). The former gave a measure of the breadth of their foreign–language knowledge, and the latter concerned the depth. Curriculum requirements meant that the majority had studied at least one language at school (usually French or German), while some had been exposed to other languages through friends, holidays, or living abroad. Full details of their second language histories are given in Appendix A.

(14)
1) None
2) I know a few words.
3) I know a few phrases and a little grammar, but not enough to use.
4) I can make myself understood in certain limited practical situations (such as shopping, asking for directions)
5) I can join in a social conversation in the language if people speak slowly and don't use difficult words.
6) I can join in a social conversation in the language.
7) I am virtually indistinguishable from a native speaker.

The trained participants had learnt a median of two foreign languages (interquartile range two to three) prior to the experiment. For the control participants the corresponding figure was 2.50 (two to three). A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and number of languages as the dependent variable did not produce a significant effect, $U = 118.500$, ns, confirming that there was no difference between the groups. Using the seven–point scale in 14, the trained participants reported a median current ability in their best foreign language of 3.75 (three to five), between *I can make myself understood in certain limited practical situations* and *I can join in a social conversation in the language if people speak slowly and don't use difficult words*. On the other hand the control participants reported a median value of 4.75 (four to five), between *I can join in a social conversation in the language if people speak slowly and don't use difficult words* and *I can join in a social conversation in the language*. A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and ability as the dependent variable produced a significant effect, $U = 77.000$, $p = 0.049$, confirming that the control participants had reached higher ability levels in their best foreign language than had the trained ones. However, if experience did affect performance, the direction of the effect would have advantaged the control rather than the trained participants. Should the trained group outperform the control therefore, it would be in spite of the difference in ability levels, and not because of it.

4.1.2 Materials

The materials in Experiment 1 were generated by the biconditional grammar shown in Figure 4.1, reproduced from Figure 2.2 (Johnstone & Shanks, 2001). As stated previously, it generates strings of eight letters divided into two halves by a full stop, using *D*, *F*, *G*, *K*, *L* and *X*. An example is shown in (15). The rules governing the strings are as follows. The letter in position one uniquely

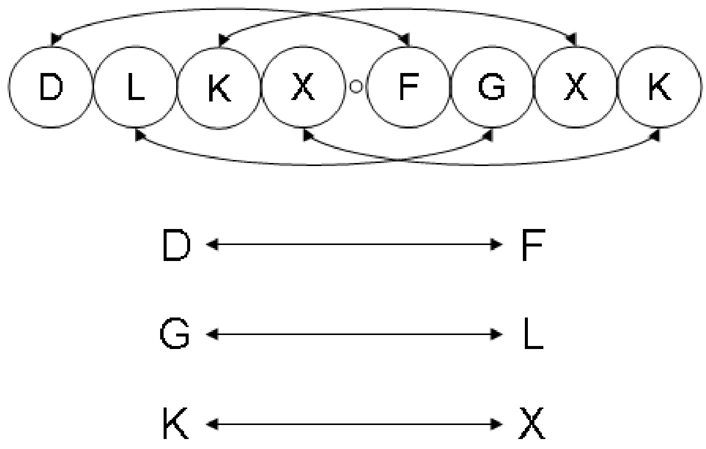


Figure 4.1: Biconditional grammar. The system produces strings of eight letters, where the first uniquely determines the fifth, the second the sixth and so on. If the first letter is *D*, the letter in the position five must be *F*. The letter pairs *G* and *L*, and *K* and *X* are similarly linked.

determines that in position five. Positions two and six, three and seven, and four and eight are linked in the same way. Thus, each letter string contains four four–step dependencies connecting positions one and five, two and six, three and seven, and four and eight. If the letter in position one is *D*, position five must be filled with *F* and *vice versa*. *G* has the same relationship with *L*, as *K* does with *X*. No immediate repetitions of a letter were permitted in the strings, with exception of repetitions between positions four and five that were separated by a full stop.

(15) DGKL.FLXG

The biconditional grammar stimuli were adapted from Johnstone and Shanks (2001, Exp 3). The original materials were used for a learning procedure called the match task, in which participants had to select the item they had just seen from a list of three. Johnstone and Shanks used some target strings more than once with different foils. These duplicates were removed, leaving twenty–four unique learning items. They are listed in Appendix D.

Two variables (grammaticality and familiarity) were manipulated in the stimuli for the subsequent judgement tests, which both employed the same forty–eight test items taken from Johnstone and Shanks (2001) (see Appendix D). The stimuli were equally divided between the four resulting conditions: grammatical–familiar (G-F), grammatical–unfamiliar (G-NF), ungrammatical–familiar (NG-F), and ungrammatical–unfamiliar (NG-NF). The example string in (15) is G-NF.

Ungrammatical items were created by breaking one of the four long–distance dependencies. Therefore, participants had to process at least as far as the fifth letter before a string could be identified as such (the exact location depended on which contingency was affected).

Familiarity was assessed in two different ways. Firstly, unfamiliar items all contained multiple novel bigrams (pairs of adjacent letters) that never appeared in any of the learning items. In contrast, all of the bigrams in the familiar items had been used in the learning items. In this way familiarity can be understood as creating a one–step dependency between adjacent letters, with the first letter in a pair determining which could follow it. The string in (15) contains six novel bigrams: *DG*, *GK*, *KL*, *FL*, *LX* and *XG*. However the remaining three, #D, LF and G# had all been used in the learning phase (# is a boundary marker representing the beginning or end of a string). In many cases as in the example string, participants did not need to read more than two letters in order to encounter a novel bigram and therefore to identify a string as unfamiliar.

The second assessment of a test string's degree of familiarity was obtained from its Associative Chunk Strength (ACS). This is a measure of the frequency of the item's constituent bigrams and trigrams. The calculation is shown in (16) for the example string, and following Johnstone and Shanks (2001). First, the frequency of each bigram and trigram in the learning phase is calculated. The ACS values for a test item are then obtained by averaging the learning phase frequencies of its constituent chunks. The onset and offset chunks that include a boundary marker as well as the adjacent letter(s) are included. The z-transformed ACS values are shown in Table 4.1 together with the values from the other experiments for comparison. The calculations were carried out independently for the timed test (based on frequency during the learning phase), the untimed test (based on the combined frequencies in the learning phase and timed test), and for the end of the experiment (based on the combined frequencies in the learning phase, timed test and untimed test). The final version (labelled post–experiment) highlights which ACS contrasts were maintained throughout the judgement tests.

(16)	Bigrams	#D	DG	GK	KL	LF	FL	LX	XG	G#
	Frequency	6	0	0	0	7	0	0	0	5
	Trigrams	#DG	DGK	GKL	KLF	LFL	FLX	LXG	XG#	
	Frequency	0	0	0	0	0	0	0	0	
	ACS	$(\frac{6+0+0+0+7+0+0+0+5}{9} + \frac{0+0+0+0+0+0+0+0}{8}) \div 2$								

		Exp 1 Biconditional	Exp 2+3 Persian	Exp 4 Basque	Exp 5 Basque
Timed	**G-F**	0.86 (0.58)	1.29 (1.21)	1.43 (0)	0.99 (0)
	G-NF	-0.93 (0.11)	-0.77 (0.49)	-0.92 (0)	-0.99 (0)
	NG-F	1.02 (0.26)	0.57 (0.46)	0.41 (0)	0.99 (0)
	NG-NF	-0.94 (0.11)	-0.02 (0.75)	-0.92 (0)	-0.99 (0)
Untimed	**G-F**	0.85 (0.57)	0.36 (1.66)	1.71 (0)	0.99 (0)
	G-NF	-0.90 (0.25)	-0.42 (0.79)	-0.57 (0)	-0.99 (0)
	NG-F	0.89 (0.71)	0.66 (0.72)	-0.57 (0)	0.99 (0)
	NG-NF	-0.84 (0.28)	-0.01 (0.82)	-0.57 (0)	-0.99 (0)
Post-Exp	**G-F**	0.76 (0.72)	1.25 (0.99)	0.19 (0)	-0.99 (0)
	G-NF	-0.83 (0.34)	-0.52 (0.83)	0.74 (0)	0.99 (0)
	NG-F	0.81 (0.88)	-0.02 (0.58)	-1.67 (0)	-0.99 (0)
	NG-NF	-0.75 (0.39)	-0.01 (0.88)	0.74 (0)	0.99 (0)

Table 4.1: Mean associative chunk strength values for each type of test item in Experiments 1 — 5. Standard deviations are shown in parentheses. The values are z-transformed to enable comparisons between experiments.

An ANOVA with grammaticality and familiarity as the independent variables and ACS in the timed test of Experiment 1 as the dependent variable did not produce a significant main effect of grammaticality, $F < 1$. However as intended there was a main effect of familiarity, $F(1, 36) = 212.606$, $p < 0.001$, $\eta_p^2 = 0.855$. Finally, there was no interaction between the variables, $F < 1$. The corresponding ANOVA for the untimed test again revealed a significant effect of familiarity, $F(1, 36) = 92.278$, $p < 0.001$, $\eta_p^2 = 0.719$, while neither the main effect of grammaticality nor the interaction reached significance, all Fs ≤ 1. A third ANOVA considered the post–experiment ACS. It confirmed that there was a significant effect of familiarity, $F(1, 36) = 49.757$, $p < 0.001$, $\eta_p^2 = 0.580$, and that there was neither a main effect of grammaticality nor a significant grammaticality x familiarity interaction, all Fs < 1. The analyses demonstrated that the set of familiar strings always had higher ACS values than the set of unfamiliar items. Additionally the analyses confirmed that the ACS values of the grammatical strings were no higher than those of the ungrammatical items. Thus, if the participants demonstrated sensitivity to grammaticality the effect could not be a consequence of sensitivity to familiarity.

Twelve NG-NF items were used for the correction test (see Appendix D). There was no way to create ungrammatical items other than by violating the target rules, so there were no fillers. Four corrections were suggested for each incorrect string. One concerned the target rule. By changing one of the letters involved in the broken long–distance dependency, it changed the item from NG-NF into G-NF. A second option increased the ACS value of the string by altering one letter to make a novel bigram familiar, although other novel bigrams remained. The final two did not improve the string.

4.1.3 Procedure

The experiment was run on PCs using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). Responses were logged either on a serial–response button box (for yes/no or multiple–choice questions) or on the computer keyboard (for the more detailed responses in the reproduction task — see Section 4.1.3 below). Participants were seated in individual sound–attenuated booths.

The trained participants were first exposed to the target structure before an immediate post–test measured whether they had acquired implicit or explicit knowledge of the rules and regularities governing the target domain. The untrained control participants took the tests without the prior learning phase. Finally, the language background questionnaires were completed on paper at the end of the session.

Learning Phase

The participants were trained using the reproduction task, variants of which have been used in many implicit learning experiments with transitional finite state grammars (A. S. Reber, 1976; A. S. Reber & Lewis, 1977; A. S. Reber, 1989). Each reproduction task trial proceeded as follows. First, a single stimulus was shown on the screen for 7000 ms (as in Johnstone & Shanks, 2001). Once it had disappeared, the participants were given unlimited time to retype it from memory using a normal keyboard. They did not need to type the full stop in the middle of the letter string. To minimise the effects of self–generated incorrect input, their production was not shown on the computer monitor and nor was there any visual confirmation that a key had been pressed. The participants were informed whether their answer was correct, but no further feedback was given. Responses were only marked as correct if they were completely accurate. When a mistake was made, participants viewed the same item again immediately and then had a second opportunity to reproduce it (with the exception of reproduction errors made on the eighth and final presentation of an item). Such repetitions did not increase the overall number of trials per item, they merely adjusted the order of presentation. Third consecutive attempts were not offered. Regardless of the success of the second reproduction, the experiment moved on and the

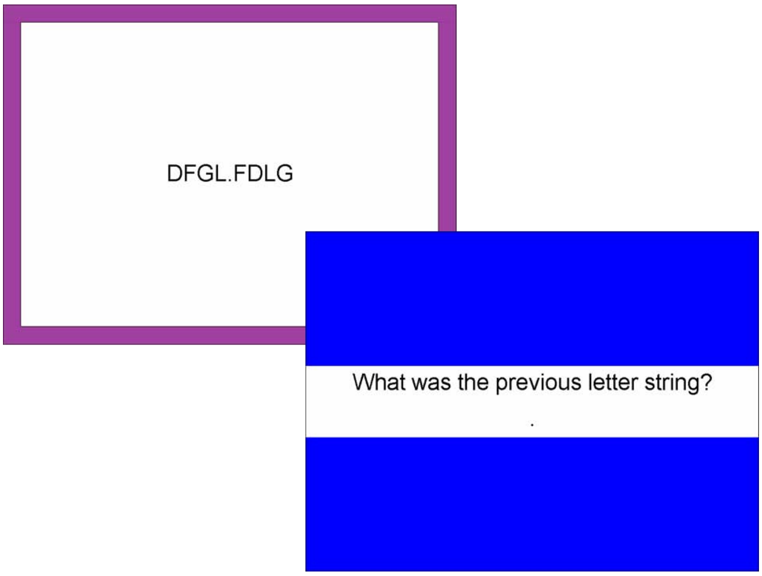


Figure 4.2: A reproduction task trial from Experiment 1.

next stimulus was presented. Figure 4.2 shows the elements that comprised a reproduction task trial, while the instructions given to the participants can be seen in Appendix B.

The exposure phase began with eight practice trials using a different letter set. This familiarised the participants with the procedure and gave them the opportunity to clarify the instructions with the experimenter. The practice trials were followed by eight trials using the individual letters that would later be employed in the experimental items (for these trials the exposure length was 2000 ms). Finally, the main learning phase began, during which each learning item was presented eight times in a random order. The participants were given the opportunity to rest at regular intervals during the procedure, but were asked to remain in the booth.

Test Phase

As discussed in Section 3.1 there were three parts to the test phase, which was based on R. Ellis (2005). They were a timed grammaticality judgement test (timed test), an untimed grammaticality judgement test (untimed test), and a multiple–choice string correction test (correction test).

The first part was the timed test in which participants had to judge whether items were correct. The stimuli were presented on the screen for 3000 ms each. The strings then disappeared before grammaticality judgements could be made. Participants were then allowed 2000 ms in which to respond (indicated by a

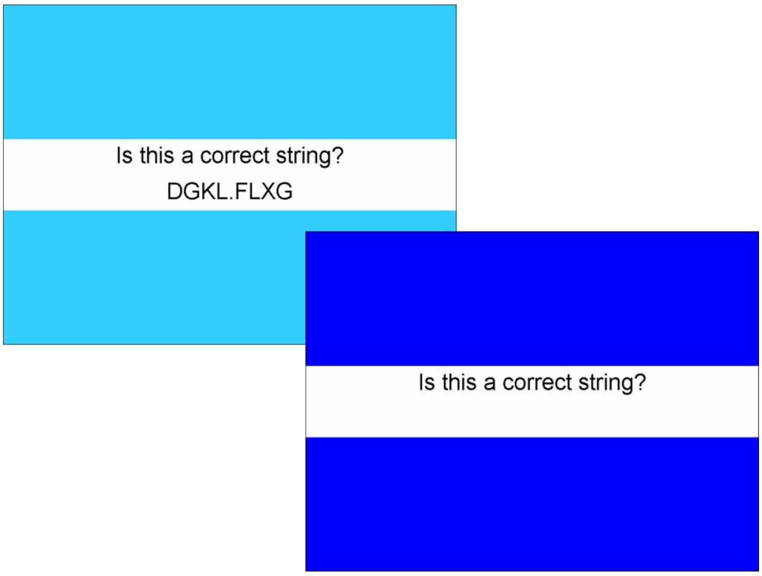


Figure 4.3: A timed grammaticality judgement test trial from Experiment 1. Responses were only permitted during the second slide, and any made earlier were not logged.

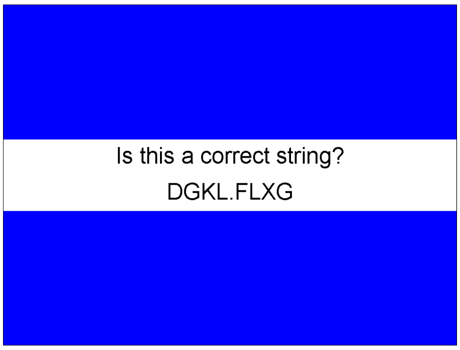


Figure 4.4: An untimed grammaticality judgement test trial from Experiment 1.

change in background colour). They pressed the leftmost key on the button–box for a judgement of *correct* and the rightmost for *incorrect*. Figure 4.3 shows a typical trial from the timed test.

The next element in the procedure was the untimed test (see Figure 4.4). Again, written versions of the stimuli appeared on the screen one at a time. Now however, they remained until a response had been logged (which was done in the same way as in the timed test). Thus, the participants could examine each item for as long as required.

The final task in the experiments was the correction test. Figure 4.5 shows the display employed in each trial. An incorrect item was presented at the top of the screen and four possible corrections were printed underneath it. The participants

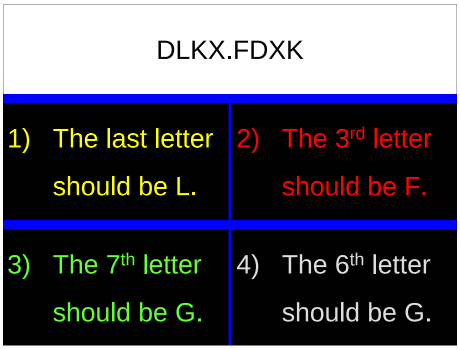


Figure 4.5: A correction test trial from Experiment 1. The test item is at the top of the screen and the suggested corrections below.

had unlimited time in which to select which of the four would successfully improve the item. Once again they responded using the button–box. The fourth button from the left was unused, and each of the remaining four was matched with one of the options using both colour–coding and numbering (with the leftmost button numbered *one* and the rightmost *four*).

4.1.4 Analysis

As described in Section 3.1, the correction test was used to divide the trained participants into three groups: rule detectors, familiarity detectors, and non-detectors. They had to answer six of the twelve correction–test questions accurately to be classed as a rule detector. If they answered the same number on the basis of familiarity they were classed as familiarity detectors. The remainder of the trained participants were labelled non-detectors. The criterion was set so that approximately 5% would be expected to reach it by chance. Appendix C shows how this was calculated.

Next, each trained group's judgement test performance was analysed against that of the control participants to investigate whether they had learnt either the target structure or the frequent bigrams. Two independent variables (grammaticality and familiarity) were manipulated in the judgement test items. Each participant's sensitivity to the two was calculated according to signal detection theory (SDT: MacMillan & Creelman, 2005). This is a statistical technique that separates a participant's sensitivity to the variables of interest (represented as d') from their overall bias towards responding 'yes' or 'no' (labelled c). The formulae, given in (17), are based on z-transformations of the number of hits (H: accurately classifying a grammatical item as correct) and false alarms (F: classifying ungrammatical items as correct). When calculating main effects of either grammaticality or familiarity it is clear what should be considered a hit and what a false alarm. However unjustified prior assumptions would be required for interactions between the two variables. For example, when calculating a grammaticality x familiarity interaction, it is unclear whether only grammatical classifications of G-F items should be considered hits, or also those of G-NF and NG-F stimuli, or indeed whether the boundary should be set in between based on the continuum of ACS values. To avoid this issue, grammaticality x familiarity interactions were not analysed.

(17) a) $d' = z(H) - z(F)$
b) $c = -{}^{1}/{}_{2}[z(H) + z(F)]$

So that performance could be assessed before and after exposure to the substantial amount of incorrect input provided by the judgement test stimuli, each test was divided into two equally–sized blocks for the analysis. The first half of the items in a test was assigned to block one, and the second half to block two. As the items were presented in a different random order for each participant, the distribution of items between blocks varied between participants.

ANOVAs with condition (rule detector or non-detector versus control) as a between–subjects factor and block (one versus two) as a within–subjects factor were used to compare the trained groups' performance (expressed in d' units) to the performance of the untrained control participants (cf. Sinnett, Costa, & Soto-Faraco, 2006). Where the ANOVAs identified a significant or marginal effect of block or a significant or marginal condition x block interaction, follow–up planned comparisons were conducted separately for each block of a judgement test using t-tests (one–tailed and with a reduced alpha level of $p < 0.025$ unless otherwise stated).

Participants were classified as non-detectors, rule detectors or familiarity detectors based on their performance in the correction test, the final part of the experiment. Whether they noticed the regularities explicitly in the learning phase or only during the judgement tests, they were still categorised as detectors. During the judgement tests therefore the rule and familiarity detector groups were a mixture of those who had already developed explicit knowledge and those who had not. The non-detector and control groups on the other hand were comparatively homogenous. This meant that, when the groups were compared statistically, the

	Control	Non-Detectors	Rule Detectors	Familiarity Detectors
Grammaticality	3.88 (0.63)	3.30 (0.39)	7.50 (1.19)	3.33 (1.45)
Familiarity	3.13 (0.45)	3.50 (0.29)	2.75 (1.18)	6.67 (0.33)

Table 4.2: Experiment 1 correction test performance. The table shows the mean (and standard error) of the number of items corrected on the basis of grammaticality (from NG-NF to G-NF) and on the basis of familiarity (by increasing the ACS) for each group of participants.

variance of the rule and familiarity detectors' performance was often significantly different to that of the controls (as demonstrated by Levene's Test). In such cases, Welch's F is reported in place of the standard F-ratio.

4.2 Results

4.2.1 Correction Test Results

Four of the sixteen trained participants reached the criterion on the correction test, answering at least six of the twelve questions by improving the string from NG-NF to G-NF. Therefore they were classified as rule detectors. Three (including one rule detector) responded to six of the twelve correction–test items on the basis of familiarity, thereby reaching the criterion to be classified as familiarity detectors. The remaining ten participants did not reach either criterion and were therefore classified as non-detectors. Table 4.2 shows the mean number of items responded to on the basis of grammaticality and familiarity for each of the three trained groups and also for the control participants.

4.2.2 Learning Phase Accuracy

Mean accuracy in the learning phase was 69.27% (standard error 4.33). In a mean of 4.72% (0.69) of the trials the correct letters were typed but in the wrong order. In the remaining 26.01% (3.86) of cases letters were omitted or incorrect letters were produced.

The learning phase performance of the three separate groups of trained participants was compared. The non-detectors correctly reproduced 66.74% of the learning items (standard error 5.54). The rule detectors responded accurately to 73.37% (9.71) of items, and the familiarity detectors to 73.73% (12.14). A one–way ANOVA with group (non-detector versus rule detector versus familiarity detector)

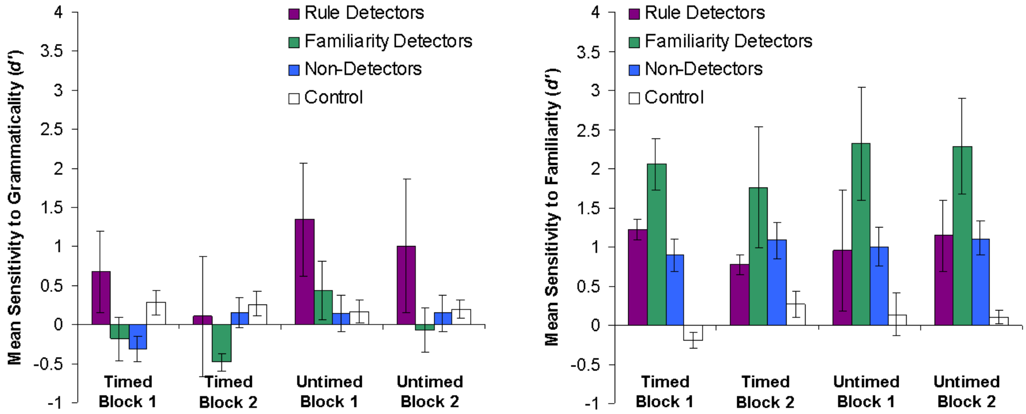


Figure 4.6: Experiment 1 results. The figure shows each group of participant's mean sensitivity to the dependent variables as measured in d' units. Error bars show the standard error.

as the independent variable and learning phase accuracy as the dependent variable did not produce a significant effect, $F < 1$. Thus, memorisation performance in the learning phase was not related to whether a trained participant became a non-detector, a rule detector or a familiarity detector.

4.2.3 *Judgement Test Results*

Each trained group's sensitivity to grammaticality and familiarity in the judgement tests (calculated according to signal detection theory) was compared to that of the control participants. Where the trained participants were significantly more sensitive to either independent variable than their control counterparts, they were assumed to have knowledge of the relevant pattern. The results are shown in Figure 4.6.

The non-detectors were expected to demonstrate implicit knowledge, as their correction test performance suggested that they had no relevant explicit knowledge with which the former could be masked. However Figure 4.6 shows that they were actually less sensitive to grammaticality than were the control participants throughout both judgement tests. Therefore it was not necessary to carry out statistical tests in order to conclude that the group had not acquired the target four–step dependencies.

Although the non-detectors did not acquire the target structure, they may have learnt the frequent bigrams. An ANOVA with condition (non-detectors versus control) as a between–subjects factor, block as a within–subject factor and sensitivity to familiarity in the timed test as the dependent variable confirmed that there was a significant main effect of condition, $F(1, 21) = 35.168$, $p < 0.001$, $\eta_p^2 = 0.626$, but neither one of block, $F(1, 21) = 2.431$, ns, $\eta_p^2 = 0.124$, nor an interaction, $F < 1$. The corresponding ANOVA for the untimed test again revealed a significant main effect of condition, $F(1, 21) = 18.643$, $p < 0.001$, $\eta_p^2 = 0.437$, and neither a main effect of block, $F < 1$, nor a condition x block interaction, $F(1, 21) = 1.205$, ns, $\eta_p^2 = 0.283$. The analyses confirmed that the non-detectors were more sensitive to the familiarity of the test strings than were their control counterparts throughout the judgement tests. The effect met two of the criteria for implicit knowledge: it was not used in the correction test and it was available in block one of the timed test. However, the third was not met as their performance was still above chance not only in block two of the timed test but also throughout the untimed test. Contrary to the prediction therefore, their knowledge of the frequent bigrams did not measurably deteriorate following exposure to unfamiliar judgement test stimuli.

The non-detectors were expected to lose the ability to distinguish familiar and unfamiliar strings when the novel bigrams became familiar during the judgement tests. This did not occur. However the z-transformed ACS values in Table 4.1 above show that the familiar items still contained more frequent bigrams and trigrams than did the unfamiliar items after the untimed test (0.87 versus -0.87), despite the use of the novel bigrams. As reported in Section 4.1.2 above, the difference was still significant. Thus the non-detectors' performance mirrored the ACS statistics, as both distinguished between the familiar and unfamiliar items throughout the judgement tests. As the group's performance mirrored the statistical distribution of the input it is possible that they relied on implicit knowledge. No firm conclusions can be drawn.

The familiarity detectors did not respond according to the target structure in the correction test. Therefore, if they were sensitive to grammaticality in the judgement tests, it could be classified as the result of implicit knowledge. However Figure 4.6 indicates that the trained participants demonstrated less sensitivity to the target structure throughout the timed test than did the controls, so no statistical tests were necessary. The ANOVA for the untimed test with condition (familiarity detectors versus controls) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable did not produce significant effects of condition or block, or an interaction, all Fs < 1. Thus, there is no evidence that the group learnt the target four–step dependencies.

Following their correction test performance, the familiarity detectors were expected to be sensitive to familiarity in the untimed test where explicit knowledge would be available, but not in the timed test. An ANOVA with condition (familiarity detector versus control) and block (one versus two) as the independent variables and sensitivity to familiarity in the timed test as the dependent variable revealed a significant effect of condition, $F(1, 15) = 43.162$, $p < 0.001$, $\eta_p^2 = 0.742$, and there was neither a significant effect of block, $F < 1$, nor an interaction, $F(1, 15) = 1.789$, ns, $\eta_p^2 = 0.201$. For the untimed test the ANOVA again showed a significant main effect of condition, $F(1, 17) = 35.843$, $p < 0.001$, $\eta_p^2 = 0.678$, but neither a main effect of block, $F < 1$, nor an interaction, $F < 1$. Their judgement test performance confirmed that the familiarity detector group had learnt the frequent bigrams. However in contrast to the prediction for explicit knowledge, their performance was above chance throughout the timed test as well as in the untimed version.

The final group of trained participants reached the criterion, answering six or more of the correction test items on the basis of grammaticality. They were expected to use the same explicit knowledge in the untimed judgement test, but not under time constraints. As predicted, an ANOVA with sensitivity to grammaticality in the timed test as the dependent variable did not produce a significant effect of condition, $F < 1$. However there was a marginal main effect of block, $F(1, 16) = 3.361$, $p = 0.085$, $\eta_p^2 = 0.174$, and a marginal interaction between the two variables, $F(1, 16) = 3.041$, $p = 0.100$, $\eta_p^2 = 0.160$. As a result, planned comparisons using one–tailed t-tests were run separately for each block. However neither produced any significant effects, block one $t(17) = 1.001$, ns, $\eta_p^2 = 0.056$, block two $t < 1$, confirming that the group was no more sensitive to grammaticality than were the control participants at any point in the timed test. In contrast the corresponding ANOVA for the untimed test indicated a significant main effect of condition, $F(1, 18) = 6.209$, $p = 0.023$, $\eta_p^2 = 0.256$, with neither the effect of block nor the interaction reaching significance, all Fs < 1. The rule detectors were only sensitive to grammaticality in the absence of time constraints and their performance was maintained in block two after the introduction of incorrect input. Thus, it met the three criteria for classification as the result of explicit knowledge.

The rule detectors demonstrated explicit knowledge of the target four–step dependencies in the correction test. However, they may also have acquired the frequent bigrams. An ANOVA with sensitivity to familiarity in the timed test as the dependent variable produced a significant effect of condition, $F(1, 16) = 32.630$, $p < 0.001$, $\eta_p^2 = 0.671$, indicating a contrast between the rule detectors and the controls. There was also a marginal condition x block interaction, $F(1, 16) = 3.819$, $p = 0.068$, $\eta_p^2 = 0.193$, but not a main effect of block, $F < 1$. A separate t-test for block one produced a main effect of condition, $t(17) = 6.524$, $p < 0.001$, $\eta_p^2 = 0.715$; however a corresponding test did not reveal a significant difference in block two, $t(17) = 1.519$, ns, $\eta_p^2 = 0.120$. In the untimed test the ANOVA again revealed a significant effect of condition, $F(1, 18) = 6.001$, $p = 0.025$, $\eta_p^2 = 0.250$, but neither the main effect of block, $F < 1$, nor the interaction reached significance, $F(1, 18) = 1.456$, ns, $\eta_p^2 = 0.075$. Thus, the rule detectors became sensitive to familiarity as well as to grammaticality, despite only using knowledge of the latter in the correction test. As expected from performance based on implicit knowledge, they were sensitive to familiarity in block one but not in block two of the timed test. However they were again throughout the untimed test. Overall therefore, their performance could not be conclusively labelled as either implicit or explicit origin.

4.2.4 Response Biases

Implicit knowledge was assumed to develop automatically and therefore to degrade following exposure to incorrect stimuli in the judgement tests. By this logic, when ungrammatical or unfamiliar items are presented sufficiently often, they should become integrated into the participants' grammar. As a result the participants should become more likely to accept other ungrammatical and unfamiliar items as correct as the experiment progresses, possibly meaning that more items are accepted overall. To test this, the response biases (c) of the non-detectors (who may have had implicit knowledge of the frequent bigrams) were calculated separately for each block of each test using the formula in (18), reproduced from (17*b*). As the group did not show any evidence of implicit knowledge of the target four–step dependencies, for the purposes of this calculation a hit was understood to be a classification of a familiar item (rather than a grammatical item) as correct. If the participants added novel bigrams to their grammar, they may have accepted more items subsequently and thereby increased the value of *c*, a possibility that was tested by means of an ANOVA.

(18) $$c = -{}^{1}/{}_{2}[z(H) + z(F)]$$

A one–way ANOVA with block (timed one, timed two, untimed one and untimed two) as the independent variable and response bias (c) as the dependent variable

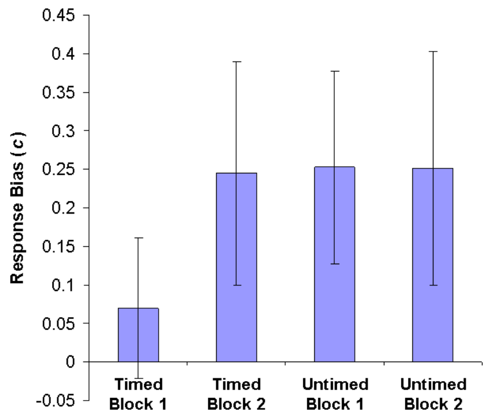


Figure 4.7: Mean response biases in Experiment 1. The non-detectors were sensitive to familiarity. Therefore the figure shows the response biases irrespective of the familiarity of an item, with responses of *correct* to familiar items considered hits and responses of *correct* to unfamiliar items labelled false alarms.

did not produce a significant effect $F < 1$. Thus, there was no evidence that the non-detectors gradually began to accept more items as correct. Nevertheless, this finding is still consistent with the claim that implicit knowledge develops automatically. The participants were likely to assume that roughly half of the items in the grammaticality judgement tests were correct and the other half were not. To maintain the balance, the threshold above which items were classified as grammatical may have been altered at the same time as the novel structures became familiar. Indeed, this assumption was made in subsequent connectionist simulations that accurately reproduced the human data (see Chapter 10).

4.3 Discussion

The findings suggested that implicit knowledge of the frequent bigrams may be acquired from the input statistics during exposure to the output of a biconditional grammar. However the results were not conclusive. The rule detector and non-detector groups were both sensitive to familiarity in the timed test and they did not respond to a sufficient proportion of correction test items according to familiarity to be familiarity detectors. However, their performance did not degrade following exposure to novel bigrams. Thus, only two of the three criteria for implicit knowledge were met. Note that although the frequent bigrams were not designated as the target structure, from the participants' point of view whether a string was familiar or unfamiliar was a perfectly valid way to represent the data.

Just as every training string was consistent with the four–step dependencies, so they all contained only frequent bigrams.

The experiment did not provide any evidence that the multiple long–distance dependencies could be acquired implicitly. Only the rule detectors were sensitive to the target structure, and their performance did not meet any of the three criteria for implicit knowledge. They reached the criterion in the correction test (by definition), their sensitivity to grammaticality could not be detected in the timed test, and their performance was maintained throughout the untimed test even following the presentation of ungrammatical input. Therefore, the rule detectors can be assumed to have used explicit knowledge of the target structure.

The present experiment confirmed that learners do not acquire implicit knowledge of the target four–step dependencies that are generated by a biconditional grammar. These results are consistent with Johnstone and Shanks (2001) who trained participants on the output of a biconditional grammar using the match task in place of the reproduction task used for the current experiment. In contrast, previous experiments using the reproduction task with shorter target structures had uncovered effects attributed to implicit knowledge (see A. S. Reber, 1989, for a review). Thus, the use of the match task alone did not prevent participants from learning the target rules in previous studies. Rather the biconditional target structure itself appears to have been responsible. Specifically, the proposal is that the increase in dependency length reduced performance.

Although implicit knowledge of the target four–step dependencies did not develop in the current experiment, the results do not prove that such knowledge can never be acquired. For example, every string generated by a biconditional grammar contains four overlapping four–step dependencies. It may be possible to acquire a four–step dependency implicitly if only one instance is found in each input string. One–step dependencies provide some evidence on this point. As discussed previously, the relationship between the two letters in a bigram can be understood as a one–step dependency: the first determines which letters can immediately follow it in a legal string. In the current experiment thirty of the forty–two possible bigrams were classified as frequent; the remaining twelve were novel. Despite the current set of materials containing thirty one–step dependencies, up to nine of which occurred in each input string, it is nevertheless possible that the non-detectors and the rule detectors acquired sensitivity to these one–step dependencies implicitly. Every AGL study that has detected implicit knowledge of bigrams has used multiple regularities in each input string (e.g.

A. S. Reber, 1989). Therefore additional regularities of an equal length may not overly complicate the input, at least where shorter patterns are concerned.

Participants have been shown to acquire implicit knowledge of shorter structures before that of longer patterns (Cleeremans & McClelland, 1991). Given this, it is unsurprising that the shorter regularities in the data may have been acquired during the relatively short experiment while the longer contingencies were not. Sensitivity to the target rules might have benefited from an increased exposure period. However, as discussed above in Section 3.4, practical considerations limited the length of the learning phase in the current experiments. One of the distinguishing features of implicit knowledge is that it can be accessed quickly. Greater practice increases the speed with which explicit knowledge can be accessed, thereby reducing the observable differences between it and implicit knowledge (DeKeyser, 2001). To avoid this problem, the issue was addressed using connectionist simulations which are reported below in Chapter 10.

Johnstone and Shanks (2001) used a recognition test in a different experiment as evidence that all of their learners' bigram knowledge was explicit. They claimed that their participants must have been using explicit knowledge when they successfully recognised bigrams from the learning phase. The recognition task has been used elsewhere as a measure of explicit knowledge (Gómez & Schvaneveldt, 1994), although it is controversial (Dienes & Perner, 1994, also see Section 2.4). Nevertheless the assumption that any test can completely exclude performance based on implicit knowledge is flawed (Jacoby, 1991). Thus, even if recognition is biased towards the use of explicit knowledge, implicit knowledge may have intruded to some extent. This problem is particularly relevant when experimenters only rely on a single criterion. In addition, Johnstone and Shanks did not use an independent measure of implicit knowledge. Showing the presence of one type of knowledge does not prove the absence of the other. To avoid this problem, the current study attempted to identify implicit knowledge directly, using three independent criteria.

4.4 Chapter Summary

The results of Experiment 1 suggest that implicit knowledge of local patterns may be acquired, consistent with claims by Gómez (1997) and A. S. Reber (1989). Previous reports that explicit knowledge is required for longer structures were also supported (e.g. Gómez, 1997). Further experiments aim to address the same question in second language acquisition, with the objective of locating the relevant length limit.

CHAPTER 5

Experiment 2

The previous experiment demonstrated that multiple one–step dependencies embedded in meaningless letter strings can be acquired and that the resulting knowledge may be implicit. However no implicit knowledge of the multiple four–step dependencies could be detected, confirming the relevance of contingency length. So far there is no evidence about input strings containing only a single instance of a regularity, which might be easier to acquire and may therefore provide a better measure of maximum performance. Similarly, although previous research has shown that adults can acquire second language structures that form one–step dependencies (N. C. Ellis, 1993; Robinson, 2005; Williams, 1999), the effect of contingency length on adult second language acquisition has not been systematically investigated.

The following experiments investigate single two– and three–step dependencies in second language acquisition. Explicit SLA clearly involves domain–general problem solving mechanisms. However, whether implicit SLA engages domain–general mechanisms or whether it relies on language–specific mechanisms is still an open empirical question. The issue is outside the range of this thesis, but the interested reader is directed to N. C. Ellis (2003) and White (2003) for contrasting opinions. If different processes are used, dependency length may not impact on the acquisition of implicit knowledge in adult second language acquisition. This conclusion would be supported if learners acquire L2 structures regardless of the contingency length. On the other hand if a comparable length limit is found to those suggested for the other paradigms, it would be consistent with a role for domain–general learning processes. Both these accounts would be compatible with the acquisition of implicit knowledge of the two–step dependency tested in

Experiment 2. However the domain–general version would not predict implicit knowledge of the longer structure, tested in Experiment 5 (for which see Chapter 8).

In Experiment 2, adult learners were trained on linguistic input with each sentence containing one two–step contingency. Participants completed the reproduction task during the learning phase. Memory tasks involving reproduction have been used in the training period of several previous implicit second language acquisition experiments (N. C. Ellis & Schmidt, 1997; Robinson, 1996, 1997a, 1997b). The use of this task also maximised comparability with Experiment 1 and with other research in the artificial grammar learning paradigm.

5.1 Method

5.1.1 Participants

Two groups of thirty–six participants were recruited in the same way as previously and the same prizes were offered. None had participated in Experiment 1. One group was trained on the target language, and the other acted as untrained controls. The participants were all monolingual native English speakers with minimal exposure to any other language before age eleven. None had studied Linguistics or English Language at university, nor had they learnt a foreign language above the level of G.C.S.E. or Standard Grade. None had any prior knowledge of the target language (Persian).

Fuller detail on the participants' language learning history was collected via a questionnaire after the experiment. The trained participants had encountered a median of two (interquartile range one to two) foreign languages prior to the experiment, while the corresponding figure for the control participants was one (one to two). A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and number of languages as the dependent variable produced a significant effect, $U = 474.000$, $p = 0.031$, indicating that there was a difference between the groups. The questionnaire also asked the participants to give self–reports of their current ability levels in their best foreign language on the seven–point scale listed above. The trained participants' median ability level was three (two to three). The median equated to *"I can make myself understood in certain practical situations"*. The same was true of the control participants. A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and ability level as the dependent variable did not uncover a significant effect of group, $U = 621.000$, ns, confirming that the two groups were equivalent in this regard.

5.1.2 Materials

The materials for Experiment 2 were based on a modified version of Persian, in which the target structure was subject–verb agreement for number. The sentences all had possessive subjects and used intransitive verbs. An example is shown in (19). There was always one intervening word (the possessor) between the subject and the verb. Assuming that each word was processed as a single element rather than as a sequence of phonemes, syllables or morphemes (as discussed above in Section 3.2), meant that the target rule was instantiated as a two–step dependency.

(19) a) Shohar-Ø-e yateem-Ø tars-e
Cousin-Ø-of doctor-Ø arrived
Subj-sg-of Possessor-sg Verb-sg
"The cousin of the doctor arrived"
b) Shohar-an-e yateem-Ø tars-ad
Cousin-s-of doctor-Ø arrived
Subj-pl-of Possessor-sg Verb-pl
"The cousins of the doctor arrived"

The experimental items were constructed from three nouns, three verbs and five grammatical affixes. The words were taken from the vocabulary lists in Alavi and Lorenz (1999), Moshiri (1988) and Samareh (1993a, 1993b) with the restrictions that the nouns all had two syllables, five or six phonemes, and six letters. The verb roots had one syllable, three or four phonemes, and four letters. Thus, even if the participants did not segment the input into word–level units, the length of the target dependency was still relatively constant across items.

Although the materials were based on a real language, several modifications were made to both the forms and the meanings. Most obviously, the written versions were transliterated into the Latin alphabet. Some sounds were also altered to clarify the grapheme–phoneme correspondences. The /ɑ/ phoneme was removed, as it is represented by the same letter as /a/ in English. Thus, *mehmaan* became *mehmun*, changing /ɑ/ to /u/ (an alternative form that is more frequent in the spoken language). The present tense verb stems were employed. There is some regularity in Persian past tense verb stems, with the majority

	Stem		Meaning	
	Original	**Modified**	**Original**	**Modified**
Nouns	Mehmaan	Mehmun	Guest	Guest
	Shohar	Shohar	Husband	Cousin
	Yateem	Yateem	Orphan	Doctor
Verbs	Geer	Geer	Receive	Ran
	Kesh	Kesh	Smoke	Left
	Tars	Tars	Fear	Arrived

Table 5.1: Persian vocabulary. The forms were modified to remove phonemes that could not be represented easily in English orthography, while the translations were altered to improve the plausibility of the items.

(including those selected for the experiment) ending in /t/ or /d/. Using the present stems therefore avoided a pattern that otherwise would have reduced the distinctiveness of the separate lexical items.

The restrictions on the form of the words made it impossible to find enough with appropriate semantics. Therefore changes were also made to the meanings used in the English translations. The nouns had to be plausible both as the possessor and as the possessee, both in the singular and in the plural. The meaning of *shohar* was changed from *husband* to *cousin* to avoid implausible phrases such as *the husbands of the doctor*. Note that the resulting form was not the actual Persian word for *cousin*, but rather that the translation used in the experiment was *cousin*. *Yateem* was translated as *doctor*, which was more suited to a possessee role than the original meaning *orphan* (avoiding instances such as *the orphans of the guest*). Common intransitive translations with relatively neutral associations were used for the verbs. Table 5.1 shows the original and modified forms of each word and of the translations.

Three nominal affixes marked singular, plural and the possessive. Two verbal affixes marked singular and plural agreement. Again, some modifications had to be made. The nominal plural morpheme *-aan*, which is used for a subset of words in Persian, was replaced by *-an*. (This removed the same phoneme as in *mehmaan*.) The verbal affixes were also altered to make them more perceptually distinct. The singular suffix *-ad* became *-e*, which is a common variant in informal speech. The plural suffix *-and* lost the nasal as the same sound was present in the plural morpheme on the noun. The reduction in euphony increased the difficulty of the learning task. The original and modified forms can be seen in Table 5.2.

	Singular		Plural		Possessive	
	Original	**Modified**	**Original**	**Modified**	**Original**	**Modified**
Nominal	Ø	Ø	-aan	-an	-e	-e
Verbal	-ad	-e	-and	-ad	N/A	N/A

Table 5.2: Persian morphology. The forms were modified to remove phonemes that could not be represented in standard English orthography, and to increase the distinctiveness of the affixes.

The lexical items and grammatical affixes listed were used to construct forty learning items, forty timed test items, forty untimed test items, and twelve correction test items (eight of which were fillers). They can all be seen in Appendix E.

Two independent variables, grammaticality and familiarity, were manipulated in the judgement test items. In each set of forty there were six G-F, fourteen G-NF, six NG-F, and the fourteen NG-NF stimuli. The imbalance between the familiar and unfamiliar items was partly caused by a difficulty in constructing the items, but also by an attempted additional manipulation which is examined in Section 5.2.4 and assessed by simulation in Section 10.2.3. The G-F items were repeated from the learning phase. Ungrammatical sentences had a number mismatch between the subject noun and the verb. Unfamiliar items had a novel possessor–verb bigram that had not been seen in the learning phase, again forming a one–step dependency.

The ACS values were included in Table 4.1 for comparison with the other experiments. In the timed test the mean z-transformed ACS value for the G-F items was 1.29 (standard deviation 1.21), for the G-NF items it was -0.77 (0.49), for the NG-F items 0.57 (0.46) and for the NG-NF items -0.02 (0.75). An ANOVA with grammaticality and familiarity as the independent variables and ACS as the dependent variable did not produce a significant effect of grammaticality, $F < 1$. However there was a significant effect of familiarity, $F(1, 36) = 28.445$, $p < 0.001$, $\eta_p^2 = 0.441$, with the familiar items having higher ACS values than the unfamiliar items. Finally there was also a grammaticality x familiarity interaction, $F(1, 36) = 8.543$, $p = 0.006$, $\eta_p^2 = 0.192$.

The ACS values were calculated again for the untimed test, based on the exposure provided in the learning phase and in the timed test and following from the assumption that learning continues during the test. In the untimed test the G-F items had a mean z-transformed ACS value of 0.36 (1.66), the G-NF items of -0.42

(0.79), the NG-F items of 0.66 (0.72) and the NG-NF items of -0.01 (0.82). The relevant ANOVA did not reveal a significant effect of grammaticality, $F(1,36) = 1.161$, ns, $\eta_p^2 = 0.031$, but there was a significant effect of familiarity, $F(1,36) = 4.798$, $p = 0.035$, $\eta_p^2 = 0.118$, with higher ACS values for the familiar than for the unfamiliar items.

Finally, the interaction between the two variables was no longer significant, $F < 1$. The post–experiment ACS values were also calculated, based on the frequencies of each bigram in the learning phase and both judgement tests combined. The mean (and standard deviation) of the ACS values for the G-F items was 1.25 (0.99), for the G-NF items it was -0.52 (0.83), for the NG-F items -0.02 (0.58), and for the NG-NF items -0.01 (0.88). The ANOVA did not produce a significant effect of grammaticality, $F(1,36) = 1.683$, ns, $\eta_p^2 = 0.045$, but there was both a significant effect of familiarity, $F(1,36) = 9.131$, $p = 0.005$, $\eta_p^2 = 0.202$, with higher ACS values for the familiar then unfamiliar items, and a significant interaction, $F(1,36) = 9.410$, $p = 0.004$, $\eta_p^2 = 0.207$.

When the experiment was planned (prior to Experiment 1), we had not recognised the possibility of learners acquiring explicit knowledge of the frequent bigrams. Therefore the test was not designed to detect any familiarity detectors. In fact, three of the four experimental items in the correction test were NG-F and only one was NG-NF. Thus it was not possible for the participants to correct all of them on the basis of familiarity. In each trial, one suggested correction addressed the number suffix on either the noun or the verb, while the three foils altered the lexical items or the word order. The eight remaining stimuli were fillers with errors irrelevant to the target structure. This reduced the chance that the participants would induce the rule explicitly during the test.

Spoken versions of the stimuli were used in the learning phase and the timed test. They were constructed from recordings of the individual words spoken in isolation by a female native speaker of Persian. The speech stream was cut into one–second segments, each containing a single word token surrounded by short silences. These tokens were concatenated without any further processing. Thus each three–word sentence lasted three seconds, and contained short pauses between each word. The resulting speech rate was very slow, but it was suitable for the participants who were absolute beginners.

5.1.3 Procedure

Like Experiment 1, the current experiment was run on PCs using E-Prime software (Schneider et al., 2002). All responses were logged either on a serial–response button box (for yes/no or multiple–choice questions) or on the computer keyboard (for the reproduction task). Participants were seated in individual sound–attenuated booths. The audio stimuli were played over headphones, with the participants able to control the volume.

As previously, the trained participants were first exposed to the target structure and then an immediate post–test measured whether they had acquired implicit or explicit knowledge of the rules and regularities governing the target domain. The untrained control participants took the tests without the prior learning phase. Language background questionnaires were completed on paper at the end of the session.

Learning Phase

During the learning phase, the participants completed the same reproduction task as in Experiment 1. In all respects other than those mentioned here, the procedure was identical in the two experiments.

Experiment 2 began with eight practice trials using German sentences. The following eight trials used the individual (uninflected) Persian words (analogous to the trials with individual letters in the previous experiment). Thus, the participants were able to learn the vocabulary to some extent before exposure to the full sentences. Newport (1988, 1990) proposed that learning this would improve learning outcomes, a claim which has since been supported both by experiments and by simulations (Elman, 1993; Kersten & Earles, 2001, but see Rohde & Plaut, 1999 for a dissenting view). These trials were followed by the main learning phase using the full Persian sentences that were each presented eight times in a random order.

As in Experiment 1, each reproduction trial began with the presentation of a learning item on the screen. However, the duration was reduced to 4700 ms as the sentences were quicker to read than the letter strings. In contrast to the previous experiment, the written stimulus was accompanied by a spoken version that was played simultaneously over headphones. The two modalities were used to prepare participants for both judgement tests, one of which only employed

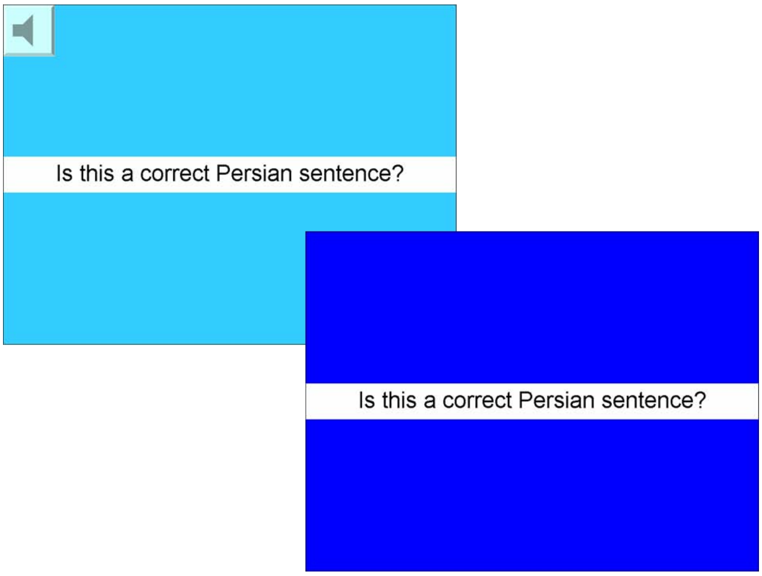


Figure 5.1: A timed grammaticality judgement test trial from Experiment 2. Responses were only permitted during the second slide, and any made earlier were not logged. Only the spoken version of the test stimulus was presented (represented by the speaker symbol).

the spoken stimuli and the other only the written (see Section 5.1.3 for more details). Once the item had disappeared, the participants were asked to retype it from memory. Their answer had to be completely accurate (including in the details of the orthography) in order to be classified as correct. If their answer was incorrect, they viewed (and heard) the item again before having a second chance to reproduce it. Third consecutive chances were not offered.

Test Phase

As in Experiment 1, the test phase started with a timed grammaticality judgement test. Whereas the previous experiment relied on written stimuli, the current one only used the spoken versions of the stimuli. The written sentences did not appear. This was a naturalistic but accurate way of controlling the exposure time. In addition, the transient nature of speech prevented the participants from returning to double–check earlier parts of the stimuli. First, a randomly selected test item was played. Once the recording had finished the background colour on the screen changed, indicating that the participants had 2000 ms in which to respond. An example trial is shown in Figure 5.1.

	Control	Non-Detectors	Rule Detectors
Grammaticality	1.17 (0.14)	1.38 (0.19)	3.50 (0.16)

Table 5.3: Experiment 2 correction test performance. The table shows the mean (and standard error) of the number of items corrected on the basis of grammaticality (from NG-NF to G-NF) for each group of participants.

The other elements in the test phase were the untimed test and the correction test. Only the written versions of the stimuli were employed in both, and therefore the procedure was identical to that in Experiment 1.

5.1.4 *Analysis*

The correction test was used to divide the trained participants into two groups: rule detectors and non-detectors. It was not possible to identify any familiarity detectors. The first group reached the criterion in the test, by answering three out of the four experimental items on the basis of grammaticality. The second did not. As previously, this criterion was set so that no more than 5% of participants could be expected to reach it by chance (see Appendix C for details of the calculation). Each group's judgement test performance was then investigated to determine the extent to which they had learnt the target structure and the frequent bigrams. This was carried out in the same way as for Experiment 1.

Note that, as the control participants only took the test phase without any prior exposure to the target language, the vocabulary as well as the grammar was novel for them. While this may have affected their performance, it would do so for all of the items equally regardless of their grammaticality or familiarity. That is, it would be unlikely to affect the measures of their sensitivity to the two variables.

5.2 Results

5.2.1 *Correction Test Results*

Twelve of the trained participants reached the criterion of answering three of the four correction test questions accurately, and they were therefore labelled rule detectors. The remaining twenty–four were classified as non-detectors. Table 5.3 summarises each group's correction test performance.

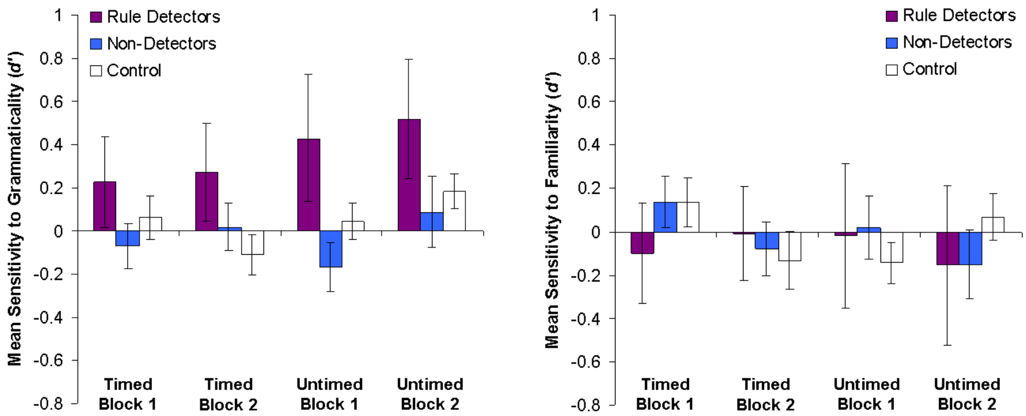


Figure 5.2: Experiment 2 results. The figure shows each group of participant's mean sensitivity to the dependent variables as measured in d' units. Error bars show the standard error.

5.2.2 Learning Phase Accuracy

During the learning phase, the trained participants completed the reproduction task and retyped every stimulus from memory immediately after it had disappeared from the screen. The overall mean performance in this task was 73.32% (standard error 2.82). The mean percentage of trials on which the response was correct *except* for the order of the letters was 0.88% (0.18) and the mean percentage of trials with other errors was 25.81% (2.74).

Learning phase accuracy was also analysed for the separate groups of trained participants. For the non-detectors the mean accuracy was 70.77% (16.57). The corresponding figure for the rule detectors was 78.41% (17.07). A two–tailed t-test with group (non-detector versus rule detector) as the dependent variable and learning phase accuracy as the dependent variable did not produce a significant effect, $t(34) = 1.290$, ns, $\eta_p^2 = 0.047$. The finding implies that learning phase accuracy was not significantly related to the development of explicit knowledge, or to the likelihood of becoming a rule detector.

5.2.3 Judgement Test Results

Figure 5.2 shows the control group's and the two trained groups' sensitivities to grammaticality and familiarity, calculated according to SDT. The non-detectors' performance provides the most direct evidence on whether any implicit knowledge of the target two–step dependency developed during the experiment. An ANOVA with sensitivity to grammaticality in the timed test as the dependent variable and condition (non-detector versus control) and block (one versus two) as independent variables did not reveal any significant main effects, all Fs < 1, nor was there a significant interaction, $F(1,58) = 1.313$, ns, $\eta_p^2 = 0.022$. No analyses were carried out for the untimed test, as Figure 5.2 shows that the trend was in the opposite direction to that predicted by the hypothesis because the untrained control participants demonstrated higher levels of sensitivity to the target structure than did the non-detectors. In summary, there is no evidence that the non-detectors had acquired the target structure.

Parallel analyses were also carried out for familiarity. The timed–test ANOVA revealed a marginal main effect of block, $F(1,58) = 3.917$, $p = 0.053$, $\eta_p^2 = 0.063$, but neither an effect of condition, $F < 1$, nor an interaction, $F < 1$. Planned comparisons performed on each block separately did not produce any significant effects, all ts < 1. Thus, the non-detectors were no more sensitive to familiarity than were the control participants at any point in the timed test. The ANOVA for the untimed test did not produce any significant effects, block x condition $F(1,58) = 2.592$, ns, $\eta_p^2 = 0.043$, all other Fs < 1. In summary, there was no evidence that the non-detectors had learnt either the frequent bigrams or the target two–step dependency. Both findings are surprising, as the contingencies are within the length–limit proposed for domain–general learning.

Figure 5.2 shows that the rule detectors were more sensitive to the target structure even in the timed test than were the control participants, but that the effect was larger once the time constraints had been removed. This pattern of results was confirmed by the analyses. An ANOVA conducted with sensitivity to grammaticality in the timed test as the dependent variable revealed a marginal effect of condition, $F(1,46) = 3.042$, $p = 0.088$, $\eta_p^2 = 0.062$, but neither a main effect of block nor an interaction, all Fs < 1. In the untimed test the effect of condition became significant, $F(1,46) = 4.166$, $p = 0.047$, $\eta_p^2 = 0.083$, and again there was neither a main effect of block nor an interaction, all Fs < 1.

An ANOVA with the rule detectors' sensitivity to familiarity in the timed test as the dependent variable did not produce a significant effect of condition, Fs < 1, of block, Fs < 1, or an interaction, $F(1,46) = 1.161$, ns, $\eta_p^2 = 0.025$. The same was true in the untimed test, condition x block $F(1,46) = 1.221$, ns, $\eta_p^2 = 0.026$, all other Fs < 1. At no point in the procedure therefore did the rule detectors demonstrate any knowledge of familiarity.

As previously there was only a small number of rule detectors, which limits the number of conclusions that can be drawn about their performance. No efforts were made to increase their numbers however, as the thesis was focused on implicit knowledge which was more likely to be identified in the performance of the non-detectors. No adjustments were made to compensate for the different numbers of participants in each group. Nevertheless, unlike the non-detectors, the rule detector group did learn the target structure and their knowledge was likely to be explicit. They reached the criterion in the correction test (by definition) and their performance did not deteriorate during the judgement tests. While they were marginally sensitive to grammaticality in the timed test suggesting that they had some access to their knowledge, the effect was only significant once the time constraints had been removed.

5.2.4 Type of Familiarity

There was a third manipulation in addition to those of grammaticality and familiarity in the Persian materials used for Experiments 2 and 3. Type of familiarity distinguished between two types of novel bigrams. One version (inflected novel) was classified as novel based on the fully inflected word forms. Thus, *shohar geere* could appear in the learning phase and *shohar geerad*, *shoharan geere* and *shoharan geerad* could all be novel bigrams in the test phase. The second type (lexically novel) could be identified based solely on the uninflected lexical items. In this case, *shohar geere* would not be accepted as novel following learning phase presentations of any of *shohar geere*, *shohar geerad*, *shoharan geere* or *shoharan geerad*. Note that difficulties designing the materials meant that the unfamiliar stimuli were not equally distributed between the two sub-types. (There were only four items with lexically novel bigrams and twenty–four with inflected novel bigrams.) If the participants were only sensitive to one type of familiarity (and particularly if they were only sensitive to the lexical version), the effect may not have been detected in the combined analyses in Section 5.2.3 above. This is one possible explanation of the null findings. Therefore the following analyses (conducted in the same way as the main ones) assessed the participants' sensitivity to each type of unfamiliarity separately. Nevertheless, as there were so few items containing lexically novel bigrams, the results must be interpreted with caution.

An ANOVA with condition (non-detectors versus control) as a between–subjects factor, block as a within–subjects factor, and sensitivity to inflected familiarity in the timed test as the dependent variable did not produce any significant effects,

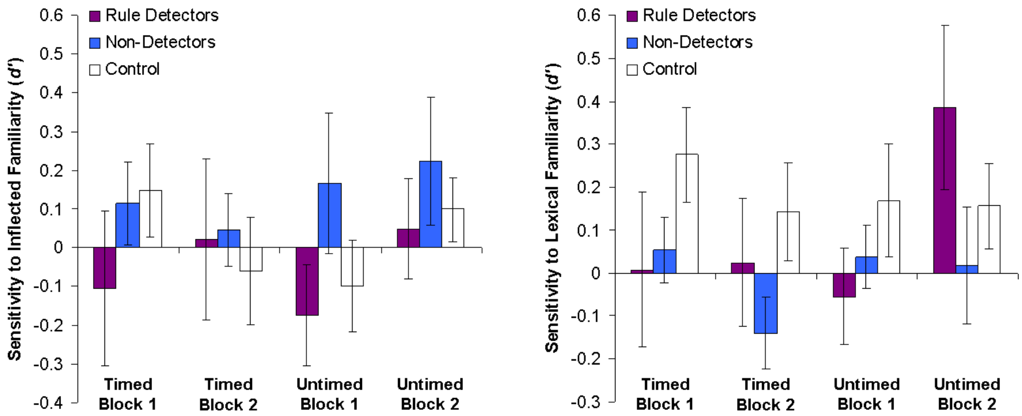


Figure 5.3: Mean sensitivity to inflected and lexical familiarity in Experiment 3. The error bars show the standard error.

block $F(1, 58) = 1.231$, ns, $\eta_p^2 = 0.021$, all other Fs < 1. The same was true for the untimed test, block $F(1, 58) = 1.119$, ns, $\eta_p^2 = 0.019$, all other Fs < 1. The non-detectors were no more sensitive to lexical familiarity than were the control participants either, with a trend in the opposite direction to that predicted in both tests (see Figure 5.3). Thus, the analyses confirmed that the non-detector group did not acquire either type of familiarity.

The rule detectors' sensitivity to both sub-types of familiarity was also assessed in the same way. In the timed test an ANOVA with sensitivity to inflected familiarity as the dependent variable did not reveal any significant effects, all Fs < 1. Figure 5.3 shows that the control participants were actually more sensitive to the manipulation in the untimed test than was the rule detector group, so inferential statistics were not required. The same was true for lexical familiarity in the timed test. The untimed–test ANOVA for lexical familiarity did not result in a significant main effect of condition, $F < 1$, or one of block, $F < 1$, but there was a marginal interaction between the two, $F(1, 35) = 4.086$, $p = 0.051$, $\eta_p^2 = 0.105$. Therefore planned comparisons were carried out using t-tests to assess each block individually. However neither produced a significant effect, block one $t < 1$, block two $t(39) = 1.061$, ns, $\eta_p^2 = 0.028$, suggesting that the rule detectors were no more sensitive to lexical familiarity than were the controls. Overall, the group did not learn either the inflected or the lexical version of the familiarity manipulation.

There was no evidence that the rule detectors or the non-detectors knew either of the types of frequent bigram. The finding is consistent with the conclusion from the main analyses in Section 5.2.3. However, as the number of items containing lexically novel bigrams was so low, it is impossible to conclusively rule out the existence of an effect.

5.2.5 Attraction Errors

The experimental stimuli all included two nouns. The subject was the first noun, the second word was another noun (the possessor), and the verb was third. As a result, an irrelevant noun intervened between the subject and the verb. Native speakers occasionally make mistakes in production (termed attraction errors) when there is a mismatch between the number values of the relevant noun and the intervening local noun (Bock & Miller, 1991; Eberhard, 1997) as do advanced second language learners (Santesteban, 2008, July). Such errors have been found both when the local noun is part of a postmodifier, as in *the key to the cabinets* (Bock & Miller, 1991), and when it is the object of the sentence (Hartsuiker, Antón-Méndez, & Zee, 2001; Santesteban, 2008, July). Research on attraction errors has focused on production, but similar effects have also been found in comprehension (Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999). Therefore performance in the judgement tests in this thesis may also have been affected.

If the learners in the current study made numerous attraction errors knowledge of the target structure may have been masked. In one half of the test items there was a match between the number values of the subject and the second noun: either both were singular or both were plural. Participants with knowledge of the target rule would be expected to respond accurately to these items. In the other half of the test items, there was a mismatch between the number values of the two nouns: if one was singular then the other was plural. Were the participants to make attraction errors to the majority of these items, then they would show a negative effect of grammaticality. Combining the two types of items as in the preceding analyses could lead to chance performance overall. In this way, attraction errors could obscure knowledge of the target rule, possibly causing the null effect in Experiment 2 (see Section 5.2.3).

Additional analyses were carried out to confirm that the main findings were not the result of attraction errors. The participants' sensitivity to grammaticality was calculated separately for those items where there was a number match and for those items with a number mismatch. The results can be seen in Figure 5.4. If the learners knew the target rule but made a large number of attraction errors, then there should be a positive effect of grammaticality in the number match items,

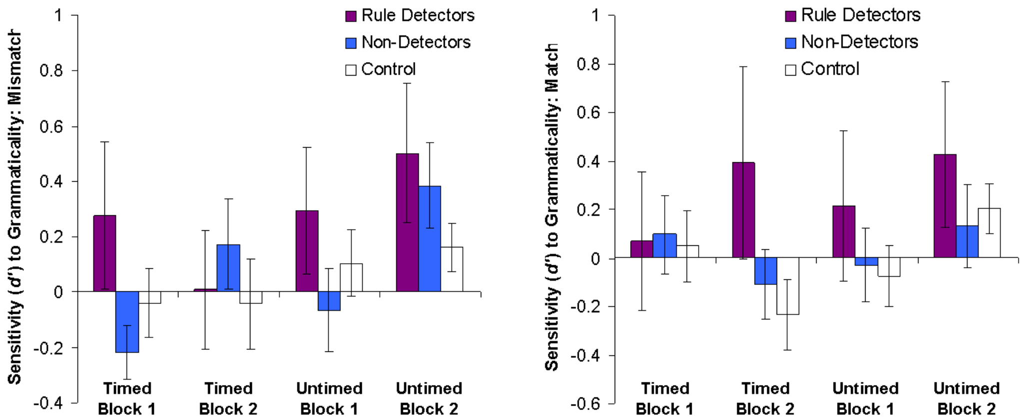


Figure 5.4: Mean sensitivity to grammaticality in Experiment 2 in the number–match and number–mismatch items as measured in d' units. The error bars show the standard error.

and depending on the proportion of attraction errors a smaller effect, a null effect, or possibly even a negative effect in the number mismatch stimuli.

As in Section 5.2.3 above, omnibus analyses were first carried out with condition and block as the independent variables and sensitivity to grammaticality as the dependent variable. When there was either a significant or a marginal effect involving block, planned comparisons using t-tests with alpha reduced to $p < 0.025$ examined each block separately. As either positive or negative effects could be expected, two–tailed t-tests were employed throughout the attraction error analyses.

Firstly, the non-detectors' performance will be discussed. In the number–mismatch items in which attraction errors were possible, an ANOVA with condition (non-detector versus control) and block as independent variables and sensitivity to grammaticality in the timed test as the dependent variable did not reveal any significant effects, condition $F < 1$, block $F(1,57) = 1.661$, ns, $\eta_p^2 = 0.028$, condition x block $F(1,57) = 1.477$, ns, $\eta_p^2 = 0.025$. In the corresponding untimed–test ANOVA there was not a significant effect of condition, $F < 1$, but there was a marginal effect of block, $F(1,55) = 3.554$, $p = 0.065$, $\eta_p^2 = 0.061$. The interaction was not significant, $F(1,55) = 1.602$, ns, $\eta_p^2 = 0.028$. Follow–up t-tests confirmed that there were no differences between the non–detectors' and the control participants' sensitivity to grammaticality anywhere in the untimed test, block one $t < 1$, block two $t(27) = 1.380$, ns, $\eta_p^2 = 0.033$.

When the same analysis was carried out on those items with a number–match between the nouns and where knowledge could not be obscured by attraction errors, the timed–test ANOVA produced a marginal effect of block, $F(1, 57) = 3.104$, $p = 0.083$, $\eta_p^2 = 0.052$, but neither the effect of condition nor the interaction reached significance, all Fs < 1. The subsequent planned comparisons on the individual blocks did not produce any significant effects of condition, all ts < 1. The corresponding ANOVA for the number–match items in the untimed test did not identify any significant effects, block $F(1, 58) = 2.696$, ns, $\eta_p^2 = 0.044$, all other Fs < 1. Overall, the non-detectors were never significantly more or less sensitive to grammaticality than were the control participants. Thus, the attraction error analyses did not provide any evidence that the non-detectors had acquired the target structure in Experiment 2.

The main analyses in Section 5.2.3 above demonstrated that the rule detector group did learn the target structure but that they could access their knowledge to a greater degree in the untimed test than in the timed version. Therefore it was important to confirm that the smaller effect in the latter was not merely the result of attraction errors. When the analysis was restricted to the number–mismatch items in the timed test, the ANOVA did not reveal any significant effects, all Fs < 1. In the corresponding untimed test ANOVA there was neither a main effect of block nor a condition x block interaction, all Fs < 1. However there was a marginal effect of condition, $F(1, 44) = 3.093$, $p = 0.086$, $\eta_p^2 = 0.066$, with the rule detector group demonstrating slightly more sensitivity to the target structure than the controls. The rule detector participants were more likely to be able to demonstrate their knowledge of the target structure in the number–match items. However the timed–test ANOVA confirmed that, even in these items, there was not a significant effect of condition, $F(1, 45) = 1.489$, ns, $\eta_p^2 = 0.032$, nor one of block, $F < 1$, nor an interaction, $F(1, 45) = 2.613$, ns, $\eta_p^2 = 0.055$. The results were no different in the untimed test, condition $F(1, 46) = 1.702$, ns, $\eta_p^2 = 0.036$, block $F(1, 46) = 2.022$, ns, $\eta_p^2 = 0.042$, condition x block interaction, $F < 1$. Thus, attraction errors do not appear to have substantially influenced the rule detectors' responses.

In summary, there was no evidence that the performance of either group of participants was substantially influenced by attraction errors. Therefore the earlier conclusions drawn in Section 5.2.3 are strengthened. The non-detectors did not acquire either the target structure or the frequent bigrams and the rule detectors were only significantly sensitive to grammaticality in the absence of time constraints.

5.3 Discussion

Implicit knowledge of the target structure was predicted to develop regardless of whether a language–specific or a domain–general learning mode was employed. However there was no evidence of any such knowledge, even when the number–mismatch items were excluded from the analysis to remove any potential attraction errors. Thus, the non-detectors did not acquire the target rule. Although the rule detectors were marginally sensitive to grammaticality under time pressure, their performance improved rather than deteriorated during the judgement tests, and they also reached the criterion in the correction test. As such, their knowledge is more likely to have been explicit. The results may imply that no long–distance dependencies can be acquired and therefore that successful use of the implicit mode in SLA is restricted to local patterns between adjacent items as learnt successfully in the experiments reported by N. C. Ellis (1993), Robinson (2005) and Williams (1999). Such a conclusion would be surprising however. Firstly, the local pattern in the current materials was not acquired either. Secondly, even though there is an upper limit for chunks and contingency lengths in other experimental paradigms, two–step dependencies and the equivalent chunks are often within that limit (Cleeremans & McClelland, 1991; Peña et al., 2002). Thirdly, language acquisition was not predicted to be more restricted than performance with other materials.

The non-detectors did not become sensitive to the frequent bigrams, despite similar regularities proving unproblematic in Experiment 1 and in previous research (A. S. Reber, 1989). Again, performance with linguistic input was below performance with artificial grammars. One possible explanation is as follows. The familiarity manipulation that created the frequent bigrams in Experiment 2 was based on lexical items rather than on parts of speech. For example, the bigram *yateem geere* was used in the learning items and *shohar tarse* was classified as a novel bigram. The two contain different lexical items, but both are noun–verb sequences. If the participants encoded the input abstractly at the level of word–classes, the two bigrams would not be distinguished. Therefore one proposed explanation is that language learners are biased towards category–based rather than lexical–item based patterns.

An alternative explanation is that the familiarity manipulation was not as strong in the Persian dataset as in the output of a biconditional grammar. The average difference between the z-transformed ACS scores of the familiar and unfamiliar items in the timed test was 1.875 in Experiment 1 and 1.325 in Experiment 2 (see Table 4.1 in Section 4.1.2 above). An ANOVA with experiment (one versus two) and familiarity as between–subjects variables and z-transformed timed ACS as the dependent variable produced a significant main effect of experiment, $F(1,84) = 3.988$, $p = 0.049$, $\eta_p^2 = 0.045$, a significant main effect of familiarity, $F(1,84) = 145.911$, $p < 0.001$, $\eta_p^2 = 0.635$, and a significant interaction, $F(1,84) = 4.282$, $p = 0.042$, $\eta_p^2 = 0.049$. The values were comparable for the familiar items. However a post-hoc t-test confirmed that the ACS values for unfamiliar items were lower in Experiment 1 than in Experiment 2, $t(50) = 3.578$, $p = 0.001$, $\eta_p^2 = 0.204$.

Not only was the ACS contrast smaller in the Persian materials, but in many cases it was necessary to take the full word forms into account to identify a novel bigram. For many of the unfamiliar items, bigrams based on the same stems as the inflected novel ones but employing different inflectional morphemes had been used in the learning phase. For example, *shohar geere* appeared in the learning phase but *shohar geerad* was still classified as a novel bigram in the judgement test items. If only the word stem was processed therefore, a supposedly novel bigram would often have been classified as frequent by the learners. (Section 5.2.4 compared this type of familiarity with that determined by the lexical items.) Given the smaller ACS contrast and the reliance on inflected familiarity, the difference between the familiar and unfamiliar items in the Persian materials may not have been learnable.

There is no evidence that either group of trained participants acquired any implicit knowledge. This may have been a consequence of the target structures, but it may also have been caused by the experimental procedure. There were more learning items in Experiment 2 than in the biconditional grammar materials used in Experiment 1 where implicit bigram knowledge may have developed. It is therefore likely that sufficient input was given for implicit learning to take place. Instead, the reproduction task itself may have been responsible. Participants had to learn the correct spellings of the modified Persian items as well as the target word order. This made the task more complex than it had been in Experiment 1, where only the order of the letters was relevant. As reported above in Section 4.2.2, there was no difference in learning phase accuracy between the experiments. Nevertheless it is still conceivable that paying attention to the orthographic detail of the stimuli may have prevented successful acquisition of the target rule. Therefore a new learning task was developed for Experiment 3 that avoided this problem. If relevant implicit knowledge was still not detected, then the target structure itself was more likely to be responsible than the task.

5.4 Chapter Summary

No implicit knowledge was detected in Experiment 2. There was no evidence that implicit knowledge of either the target two–step dependency or the one–step familiarity manipulation was acquired. This was contrary to the predictions, as similar structures are acquired in other experimental paradigms (Cleeremans & McClelland, 1991; Peña et al., 2002; A. S. Reber, 1989). As the result was unexpected, Experiment 3 tested whether the same linguistic target structures can be acquired during a different learning task that requires less attention to detail.

Chapter 6

Experiment 3

Unexpectedly there was no evidence that the participants in Experiment 2 acquired implicit knowledge of either the target two–step dependency or the frequent bigrams. One possible explanation is that the reproduction task prevented measurable implicit knowledge from developing when the participants also needed to learn how to spell each word. Therefore a new learning task (source localisation) was introduced for use with the same materials in Experiment 3. In all other respects, Experiment 3 was identical to Experiment 2.

The source localisation task was inspired by the match task, a procedure that has been used to train participants in experiments with biconditional grammars. For example, Johnstone and Shanks (2001) presented participants with individual strings for seven seconds each. After a two second pause a set of three strings was displayed and participants had to identify the one they had recently seen. The two foils were illegal versions of the target string. Thus, participants only had to remember a subset of the training items at any one time.

The match task was felt to be an improvement on the reproduction task for experiments into second language acquisition because the participants did not need to learn the orthography in order to answer correctly. Nevertheless, we were concerned that 50% of the input that participants received during the original version of the task was actually incorrect. Recall that, as argued in Section 2.1, implicit knowledge is acquired automatically regardless of the grammaticality of the input. Therefore such a large proportion of incorrect input could prevent learners from acquiring implicit knowledge that is representative of the target structure (Dienes et al., 1991; Gómez & Lakusta, 2004).

The source localisation task was an adaptation of the match task, designed to eliminate exposure to incorrect input and to avoid the requirement to learn the orthography. In addition it allows participants to benefit from additional sources of information available to language learners. Specifically, English translations were included to provide semantics. The exact discrimination that participants made was adjusted so that they had to hold multiple items in memory simultaneously, a procedure that may have increased the likelihood of generalisations.

6.1 Method

6.1.1 Participants

Thirty–six trained participants were recruited in the same way and from the same population as in the previous experiment. None had taken part in the previous experiments. No additional control participants were recruited; instead, the control data from Experiment 2 were reused. The control participants only took part in the test phase, during which the two experiments were identical.

Examination of the language background questionnaires indicated that the median number of languages the trained participants had encountered prior to the experiment was two (interquartile range one to two). (For the control participants the corresponding figure was one (one to two), see Section 5.1.1 above.) A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and number of languages as the dependent variable was used to compare the new trained participants with the control participants from Experiment 2 in this regard. It did not indicate a significant effect, $U = 526.000$, ns. The participants' current ability in their best foreign language was also reported on the same seven–point scale as previously. The median ability level for the trained participants was three (two to three). The median translated as *I can make myself understood in certain limited practical situations*. The same was true for the control participants. A two–tailed Mann–Whitney test with group as the independent variable and ability as the dependent variable did not reveal a significant effect, $U = 646.500$, ns. Overall therefore, the trained and control participants had comparable amounts of experience learning foreign languages prior to the experiment.

6.1.2 Materials

The materials in Experiment 3 were identical to those employed in Experiment 2. Again, the target structure was subject–verb agreement in a modified version of Persian, instantiated as a two–step dependency.

6.1.3 Procedure

As previously, the experiment was run on PCs using E-Prime software (Schneider et al., 2002). Participants sat in individual sound–attenuated booths. The audio stimuli were played over headphones, with the participants able to control the volume. As all the questions were now either yes/no or multiple–choice the keyboard was no longer required, and all responses were logged on a serial–response button box.

Learning Phase

As in Experiment 2 the experiment began with eight practice trials in German, followed by another eight which introduced the individual (uninflected) Persian words. Finally the main learning phase began, with three hundred and twenty trials (eight repetitions of each item) using Persian sentences.

The participants in Experiment 3 carried out a new task in the learning phase: source localisation. Crucially, this task did not involve exposure to incorrect input. Appendix B shows the instructions given to the participants. Each trial proceeded as follows. It began with a picture of one of four speakers that appeared on the screen for 700 ms. This was replaced by a learning stimulus that was displayed for 4700 ms while the spoken version was played over headphones (the exposure length was reduced to 2000 ms for the trials with the individual words). Regardless of which speaker was pictured, the same voice was used. Second language learners at lower proficiency levels tend to develop speaker–specific representations, so this aspect of the procedure eliminates a factor that may otherwise have impeded performance (Trofimovich, 2005). After the target item, a written English translation was displayed on the screen for 1500 ms. Including the spoken versions and the translations ensured that the linguistic stimuli included both phonetics and semantics. Figure 6.1 shows the elements of a source localisation trial.

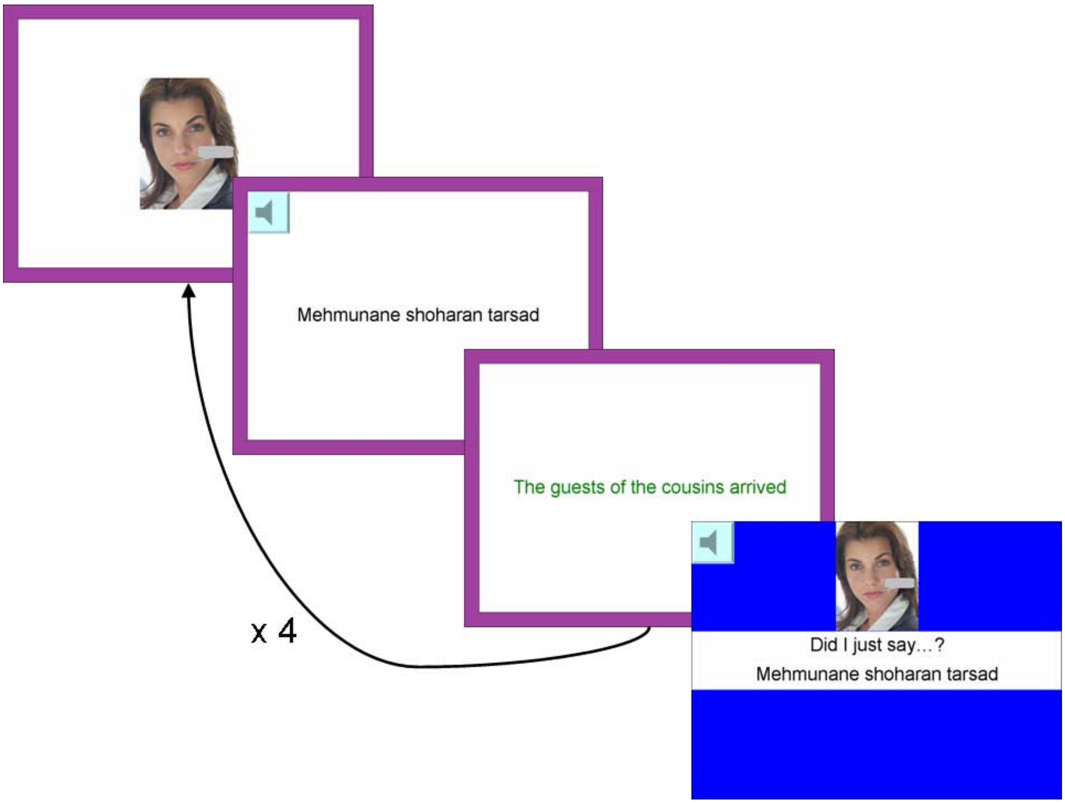


Figure 6.1: A source localisation task trial. The speaker symbol represents the spoken stimuli. Participants were presented with four speaker–sentence pairs and were then asked a localisation question.

After four such trials (one with each picture), participants answered a source localisation question without any time restriction. One of the pictures was displayed on the screen with a target sentence and the question 'Did I just say...?' They were instructed to reply 'yes' by pressing the rightmost button on the button–box if the speaker had been linked to that sentence in the previous block of four trials (as was the case in half of the items), and 'no' by pressing the leftmost button otherwise (including if the speaker and sentence had been paired earlier in the experiment). Thus, it was necessary to hold a maximum of four sentences in memory at any one time. No feedback was given. Unbeknownst to the participants, the foils were grammatical sentences selected from the set of learning items that had *not* been employed in the previous block of four trials. Thus, it was possible to complete the task accurately just based on memory of the preceding four sentences rather than of the sentence–picture pairs and without detailed knowledge of the orthography.

	Control	Non-Detectors	Rule Detectors
Grammaticality	1.17 (0.14)	1.17 (0.11)	3.46 (0.14)

Table 6.1: Experiment 3 correction test performance. The table shows the mean (and standard error) of the number of items corrected on the basis of grammaticality (from NG-NF to G-NF) for each group of participants.

Test Phase

The procedure in the test phase was identical to that in Experiment 2 in all respects.

6.1.4 *Analysis*

As in Experiment 2, the correction test was used to divide the trained participants into rule detectors and non-detectors. Again it was not possible to identify any familiarity detectors, because the test did not allow participants to improve the items by increasing their ACS value. The first group reached the criterion in the test, answering three or more of the four experimental items on the basis of grammaticality. The second group did not. Each group's judgement test performance was then contrasted with that of the same control participants as in Experiment 2.

6.2 Results

6.2.1 *Correction Test Results*

The correction test responses were analysed first in order to classify the trained participants as either non-detectors or as rule detectors. Thirteen of the thirty–six trained participants were rule detectors, with the remaining twenty–three non-detectors not reaching the criterion. This figure is very similar to Experiment 2 where twelve participants became rule detectors following training on the reproduction task with the same materials. Thus, the two learning tasks were likely to be equivalent in terms of the amount of explicit knowledge they engendered. Table 6.1 shows the mean number of items answered correctly by the two trained groups and by the control participants.

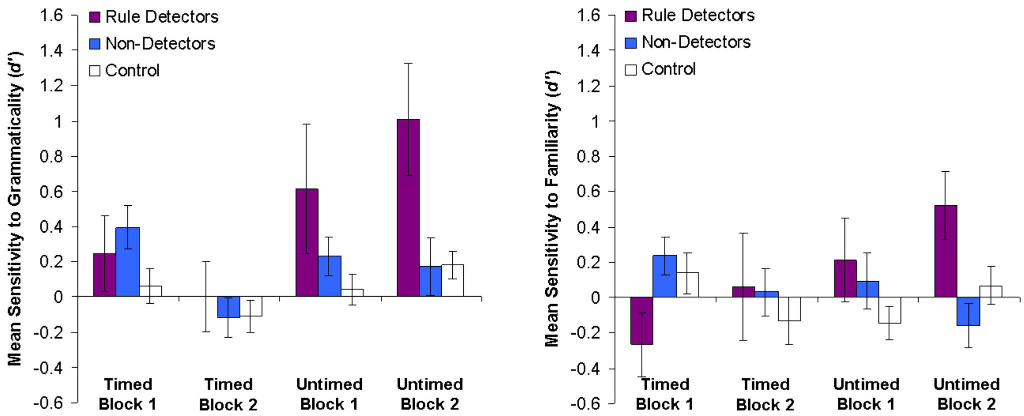


Figure 6.2: Experiment 3 results. The figure shows each group of participant's mean sensitivity to the dependent variables as measured in d' units. Error bars show the standard error.

6.2.2 *Learning Phase Accuracy*

The trained participants in Experiment 3 completed the source localisation task, rather than the reproduction task used in the earlier experiments. In the learning phase the participants responded to a mean of 69.27% of the source localisation questions accurately (standard error 2.89). The non-detectors' and rule detectors' accuracy in the learning phase was also analysed separately. The former group's mean accuracy was 67.82% (8.79), while the latter's was 74.25% (11.42). A two–tailed t-test with group (non-detector versus rule detector) as the independent variable and learning phase accuracy as the dependent variable produced a marginal effect, $t(34) = 1.891$, $p = 0.067$, $\eta_p^2 = 0.095$. Thus, the rule detectors may have responded correctly to a greater proportion of the source localisation questions than did the non-detectors, and therefore they may have had a better memory for the input forms.

6.2.3 *Judgement Test Results*

Figure 6.2 shows each group's sensitivity to grammaticality and familiarity, expressed in d' units as previously. Following the change in learning task, the non-detectors were expected to acquire measurable implicit knowledge. An ANOVA with condition (non-detectors versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality in the timed test as the dependent variable revealed a marginal effect of condition, $F(1, 57) = 2.789$, $p = 0.100$, $\eta_p^2 = 0.047$, and a significant effect of block, $F(1, 57) = 8.266$, $p = 0.006$, $\eta_p^2 = 0.127$, but not an interaction between the two, $F(1, 57) = 2.029$, ns, $\eta_p^2 = 0.034$.

One–tailed t-tests for the individual blocks produced a significant effect of condition in block one, $t(57) = 2.076$, $p = 0.021$, $\eta_p^2 = 0.070$, but not in block two, $t < 1$. The corresponding ANOVA for the untimed test did not produce any significant effects, all Fs < 1. The analyses suggested that the non-detectors had acquired the target two–step dependency, and they had done so according to the three criteria for implicit knowledge. Their knowledge was not employed in the correction test, and their performance was above that of the controls in block one of the timed test but not in block two. In contrast to the reproduction task therefore, source localisation appeared to allow second language learners to develop implicit knowledge of the regularities in the current materials.

As the non-detectors acquired implicit knowledge of the target structure, they were also expected to acquire the familiar bigrams, a shorter structure. However an ANOVA with sensitivity to familiarity in the timed test as the dependent variable did not reveal a significant effect of condition nor a condition x block interaction, all Fs < 1, although there was a marginal effect of block, $F(1,57) = 3.744$, $p = 0.058$, $\eta_p^2 = 0.062$. The resulting planned comparisons did not produce a significant effect of condition for either block, all ts < 1. Contrary to the prediction, the non-detectors were not sensitive to familiarity at any point in the test. The corresponding untimed–test ANOVA did not produce significant main effects of either condition or block, all Fs < 1, but there was a marginal condition x block interaction, $F(1,57) = 3.076$, $p = 0.085$, $\eta_p^2 = 0.051$. Planned comparisons did not reveal a significant effect of condition in either block, all ts < 1, thus confirming that the non-detectors were not sensitive to the frequent bigrams in the untimed test either. As the participants did not learn the frequent bigrams in the Persian materials in either Experiment 2 or Experiment 3, it is likely that the pattern itself was not learnable. However, the current data do not distinguish between the potential explanations proposed in Section 5.3 above, namely the small ACS contrast and the reliance on inflected familiarity. Instead, this will be assessed using simulations in Section 10.2.3.

The rule detectors' performance in the correction test suggested that they had explicit knowledge of the target structure. However they were not expected to be able to access this knowledge under time constraints. An ANOVA for the timed test confirmed that there was no significant main effect of condition, $F(1,47) = 1.118$, ns, $\eta_p^2 = 0.023$, nor a main effect of block, $F(1,47) = 1.953$, ns, $\eta_p^2 = 0.040$, nor an interaction, $F < 1$. Figure 6.2 shows that there was marked difference between the rule detector group's performance in the timed and untimed tests, supporting the assumption that explicit knowledge can only be accessed without time pressure. When the time constraints were removed, the untimed test ANOVA revealed the predicted main effect of condition, $F(1,47) = 14.018$, $p < 0.001$, $\eta_p^2 = 0.230$, but not one of block, $F(1,47) = 2.685$, ns, $\eta_p^2 = 0.054$, nor an interaction, $F < 1$. The effect did not fit any of the three criteria for performance based on implicit knowledge, and so can be labelled as explicit in origin. The group performed at above–chance levels in the correction test but not in the timed test, and their untimed test performance was maintained throughout both blocks. The rule detectors had learnt the target structure, but unlike the non-detectors whose knowledge could only be detected at the beginning of the judgement tests, their performance improved towards the end.

The rule detectors' knowledge of the bigrams was also assessed. An ANOVA with sensitivity to familiarity in the timed test as the dependent variable confirmed that there were no main effects of condition or of block, all Fs < 1, and nor was there an interaction, $F(1,47) = 2.680$, ns, $\eta_p^2 = 0.054$. On the other hand an ANOVA for the untimed test produced a significant main effect of condition, $F(1,47) = 8.027$, $p = 0.007$, $\eta_p^2 = 0.146$, and a marginal one of block, $F(1,47) = 2.818$, $p = 0.100$, $\eta_p^2 = 0.057$, but again no interaction, $F < 1$. According to the subsequent planned comparisons, the rule detectors were marginally more sensitive to familiarity in block one than were the control participants $t(47) = 1.697$, $p = 0.048$, $\eta_p^2 = 0.058$, and they were significantly so in block two $t(47) = 2.128$, $p = 0.020$, $\eta_p^2 = 0.088$. This pattern, where performance did not degrade between blocks and was also not above chance in the timed test, is suggestive of explicit learning. However, as there was no measure of explicit bigram knowledge in the correction test, the final criterion could not be assessed.

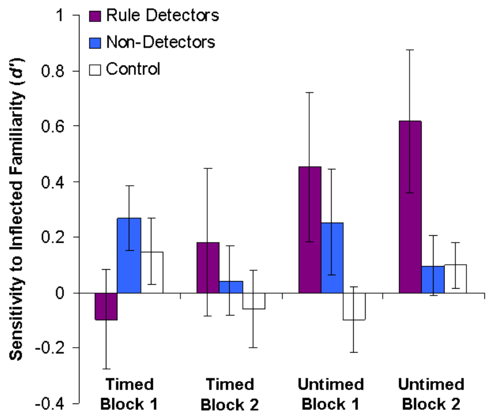
6.2.4 *Type of Familiarity*

As in Experiment 2 the participants' sensitivity to each type of familiarity was analysed separately. If the learners were able to discriminate only between lexically familiar and novel bigrams or only between inflected familiar and novel bigrams, the combined analyses in the previous section may not have been able to reveal the effect. However, as only four items contained lexically novel bigrams, the results must be interpreted with extreme caution.

An ANOVA with condition (non-detector versus control) and block (one versus two) as the independent variables and sensitivity to inflected familiarity in the timed test as the dependent variable did not reveal a significant main effect of condition nor a condition x block interaction, all Fs < 1. However as there was a marginal main effect of block, $F(1, 57) = 3.207$, $p = 0.079$, $\eta_p^2 = 0.053$, planned comparisons were conducted for each one individually. The one–tailed t-tests did not produce any significant differences between the groups in either block, all ts < 1. The corresponding ANOVA for the untimed test did not uncover any significant effects, condition $F(1, 57) = 2.028$, ns, $\eta_p^2 = 0.034$, block $F < 1$, condition x block $F(1, 57) = 1.896$, ns, $\eta_p^2 = 0.032$. Overall therefore there was no evidence that the non-detectors had acquired the inflected familiar bigrams.

Although the non-detectors were not sensitive to inflected familiarity, they may still have learnt the lexical familiarity pattern. An ANOVA for the timed test did not produce a significant effect of condition, $F(1, 46) = 1.900$, ns, $\eta_p^2 = 0.040$, but there was a marginal effect of block, $F(1, 46) = 3.027$, $p = 0.089$, $\eta_p^2 = 0.062$, and a significant interaction, $F(1, 46) = 6.719$, $p = 0.013$, $\eta_p^2 = 0.127$. The resulting planned comparisons did not identify a significant contrast between the non-detectors and the controls in block one, $t < 1$, but the non-detectors were marginally more sensitive to lexical familiarity than were the control participants in block two, $t(54) = 1.785$, $p = 0.040$, $\eta_p^2 = 0.056$. The parallel ANOVA for the untimed test did not reveal any significant main effects, condition $F < 1$, block $F(1, 45) = 1.212$, ns, $\eta_p^2 = 0.026$, but there was a significant interaction, $F(1, 45) = 4.513$, $p = 0.039$, $\eta_p^2 = 0.091$. The planned comparisons demonstrated that the non-detectors and the control participants did not differ significantly in the untimed test with regard to lexical familiarity, block one $t(51) = 1.143$, ns, $\eta_p^2 = 0.025$, block two $t < 1$. In summary, the non-detectors' performance in the timed test suggested some knowledge of lexical familiarity, but as it was only apparent in block two of the timed test, it did not meet the requirements to be classified as implicit. Given the small number of items, the result could be spurious.

Parallel analyses were carried out for the rule detectors. The timed–test ANOVA for inflected familiarity did not produce any significant main effects, all Fs < 1, nor a condition x block interaction $F(1, 47) = 1.833$, ns, $\eta_p^2 = 0.038$. In contrast, the corresponding untimed–test ANOVA revealed a significant main effect of condition, $F(1, 47) = 8.753$, $p = 0.005$, $\eta_p^2 = 0.157$, but not one of block, $F(1, 47) = 1.596$, ns, $\eta_p^2 = 0.033$, or an interaction between the two, $F < 1$. Thus, the rule detectors were sensitive to inflected familiarity. As their performance was at chance in the timed test and it did not deteriorate between blocks one and two



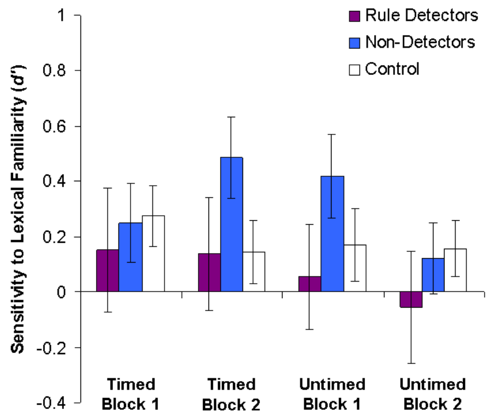


Figure 6.3: Mean sensitivity to inflected and lexical familiarity in Experiment 3 as measured in d' units. The error bars show the standard error.

of the untimed test it was likely to be based on explicit knowledge. Remember however that the correction test items in Experiment 3 did not allow participants to respond on the basis of familiarity, so the third criterion, that of also being a familiarity detector, could not be assessed.

The rule detectors appeared to develop explicit knowledge of the inflected familiar bigrams, which like the target structure involved the grammatical morphemes. In contrast however, Figure 6.3 demonstrates that they were never more sensitive to lexical familiarity than were the control participants and therefore there was no evidence that the group had learnt this pattern.

Although the rule detectors acquired the inflected–familiar bigrams, their knowledge was probably explicit. There was no evidence that either group acquired implicit knowledge of either type of familiar bigram (although this could be due to the low number of items for the lexical version). However, as this result was similar to that in Experiment 2, we can conclude that the materials were more likely to have caused the difficulties than was the learning task.

6.2.5 *Response Biases*

In Experiment 3 the non-detectors acquired implicit knowledge of the target structure. Their performance met all three of the criteria: by definition they did not reach the criterion in the correction test, they were also above chance in the timed judgement test, and their performance was less accurate in block two than in block one. Specifically, the non-detectors' knowledge appeared to become less representative of the target structure when they received exposure to incorrect

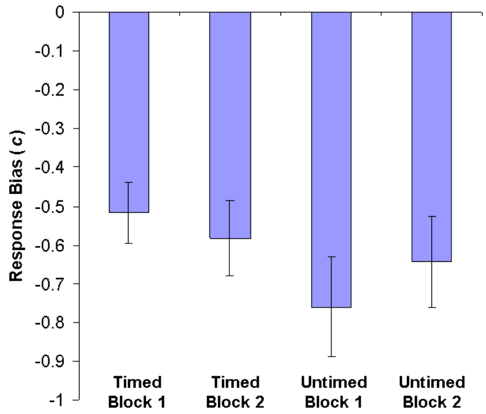


Figure 6.4: Mean response biases in Experiment 3. The non-detectors were sensitive to grammaticality. Therefore the figure shows the response biases irrespective of the grammaticality of an item, with responses of *correct* to grammatical items considered hits and responses of *correct* to ungrammatical items labelled false alarms. The error bars show the standard error.

input in the judgement tests. If the new ungrammatical input was integrated into the grammar, then the participants may have gradually begun to accept a greater proportion of the items as correct. Such an adjustment would lead to an increase in the response bias (*c*). This statistic, calculated from the numbers of hits and false alarms according to the formula given previously in (18) measures the likelihood of a participant classifying an item as correct regardless of whether it is actually grammatical.

Figure 7.2 shows the mean values of *c* in each judgement test block. If anything, it suggests that the participants became less likely to accept test strings as the experiment progressed. A one–way ANOVA with block (timed one, timed two, untimed one, untimed two) as the independent variable and *c* as the dependent variable tested whether the trend was significant. The analysis did not produce a significant effect of block, $F(3, 66) = 1.352$, ns, $\eta_p^2 = 0.058$. Therefore the non-detectors did not gradually accept more or indeed fewer items as grammatical following exposure to incorrect judgement test items. Instead, if they integrated the incorrect items into their representation of the grammar, they must also have adjusted the threshold above which items were classified as grammatical. (The same claim was made for Experiment 1 in Section 4.2.4 above.)

6.2.6 Attraction Errors

As in Experiment 2, the participants' responses may have been affected by attraction errors. The claim that the non-detectors had implicit knowledge partly depended on their becoming less sensitive to grammaticality by block two of the timed test. However, if this finding were solely the result of a growth in attraction errors to the number–mismatch items, the conclusion could not be upheld.

There is also an additional reason to examine the attraction error analyses for Experiment 3. Again, as a result of difficulties when designing the stimuli the two possessor and the subject had the same number value in twenty-eight of the forty learning items, and there was only a mismatch in the remaining twelve. This imbalance may have led participants to abstract the erroneous generalisation that the number of the verb was dependent on the number of the second noun (the possessor) rather than of the first (the possessee). Thus, they may have acquired a one–step rather than a two–step dependency. A similar imbalance in the test materials meant that such knowledge was likely to result in apparent sensitivity to familiarity. Although there was no evidence for this effect in the main analyses, a more complete picture was obtained by examining the match and mismatch items separately in the attraction error analyses. If participants had acquired a two–step dependency they should be sensitive to grammaticality with both types of item. However, if they had acquired the one–step contingency then they should only be more sensitive to grammaticality than the controls in the match items and less so in the mismatch items.

As in Section 5.2.5 above, the participants' sensitivity to grammaticality was calculated separately for items with a number match between the two nouns (where attraction errors were not possible) and for those items with a number mismatch (where attraction errors could occur). The values are shown in Figure 6.5. Separate ANOVAs were used for the number match and mismatch items, with condition and block as the independent variables and sensitivity to grammaticality as the dependent variable. If an ANOVA produced a significant or marginal effect involving block, it was followed up with two–tailed t-tests performed separately for each block, with alpha reduced to $p < 0.025$.

As usual, the non-detectors' performance will be reported first. An ANOVA considering only the number–mismatch items in the timed test with condition (non-detector versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable did not

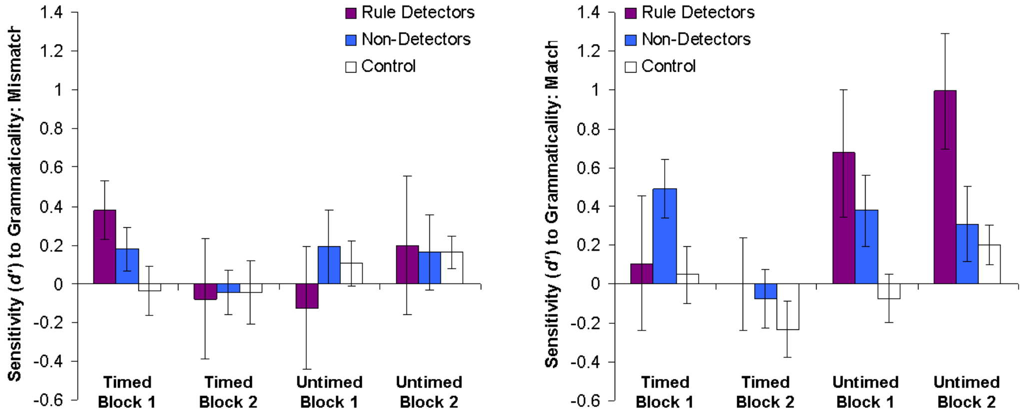


Figure 6.5: Mean sensitivity to grammaticality in Experiment 3 in the number–match and number–mismatch items, as measured in d' units. The error bars show the standard error.

produce any significant effects, all Fs < 1. The same was true for a corresponding ANOVA in the untimed test, all Fs < 1. In contrast, the results for the number–match items paralleled those for the whole data set reported in Section 6.2.3 above. The timed–test ANOVA revealed a significant effect of block, $F(1,55) = 12.751$, $p = 0.001$, $\eta_p^2 = 0.188$, but neither one of condition, $F(1,55) = 2.494$, ns, $\eta_p^2 = 0.043$, nor an interaction, $F(1,55) = 1.892$, ns, $\eta_p^2 = 0.033$. Planned comparisons between the non-detector and the control participants were carried out for each block separately. The t-tests indicated that the non-detectors were marginally sensitive to grammaticality in block one of the timed test, $t(57) = 2.005$, $p = 0.050$, $\eta_p^2 = 0.066$, but not in block two, $t < 1$. In the untimed test on the other hand an ANOVA did not reveal a significant main effect of condition, $F(1,56) = 2.022$, ns, $\eta_p^2 = 0.035$, a main effect of block, $F < 1$, or an interaction, $F(1,56) = 1.887$, ns, $\eta_p^2 = 0.033$. As the non-detectors' sensitivity to grammaticality could only be detected in the number–match items, attraction errors may have reduced the effect for the mismatch stimuli. Importantly however, they were less sensitive to grammaticality in the mismatch items than were the control participants, suggesting that they had acquired the target two–step and not the confounding one–step dependency. Overall therefore the main pattern of results and conclusions presented in Section 6.2.3 above was not qualitatively altered.

Equally, it was important to confirm that attraction errors had not obscured any sensitivity to grammaticality in the timed test for the rule detectors. An ANOVA examining sensitivity to grammaticality in the number–mismatch stimuli in the timed test did not produce a significant effect of condition, $F(1,46) = 1.043$, ns, $\eta_p^2 = 0.022$, of block, $F < 1$, or a condition x block interaction, $F < 1$. In contrast to the total data set analysed in Section 6.2.3 above, the corresponding ANOVA did not reveal any significant effects in the untimed test either, block $F(1, 45) = 1.278$, ns, $\eta_p^2 = 0.028$, all other Fs < 1. Next, performance on the number–match items was analysed. An ANOVA did not identify any significant effects for the rule detectors in the timed test, all Fs < 1. In the untimed number–match items ANOVA on the other hand, there was a significant effect of condition, $F(1, 47) = 18.221, p < 0.001, \eta_p^2 = 0.279$, a marginal effect of block, $F(1, 47) = 3.730, p = 0.059$, $\eta_p^2 = 0.074$, and no interaction, $F < 1$. Planned comparisons confirmed that the rule detectors were significantly more sensitive to grammaticality in the number–match items than were the control participants both in block one of the untimed test, $t(48) = 2.483, p = 0.025, \eta_p^2 = 0.163$, and in block two, $t(48) = 2.865, p = 0.012$, $\eta_p^2 = 0.227$.

The main analyses suggested that both groups of trained participants in Experiment 3 learnt the target structure. The analyses in this section supported that claim, confirming that the participants had not acquired the confounding one–step dependency between the number of the possessor and the number of the verb. In addition, the analyses indicated that they could only respond accurately to the number–match items where attraction errors were impossible. When attraction errors could occur in the number–mismatch items, both groups of trained participants made sufficient errors to obscure their knowledge of the subject–verb agreement. Despite this, the proportion of attraction errors was not high enough to qualitatively alter the overall pattern of results obtained in the combined analyses in Section 6.2.3. In no cases were the participants significantly sensitive to grammaticality in the number match items without also being so when the whole data set was included in the analysis. Therefore the earlier conclusions remain unchallenged.

6.3 Discussion

During the source localisation task the participants were able to acquire the target two–step dependency implicitly. The non-detectors were sensitive to grammaticality and their responses met the three criteria of performance resulting from implicit knowledge. This was expected regardless of whether or not a length limit exists for the acquisition of implicit knowledge of linguistic structures. Therefore, Experiment 5 examined three–step dependencies. Structures of this length are unlikely to be amenable to implicit learning or the acquisition of implicit knowledge in other experimental paradigms (Cleeremans & McClelland, 1991; Johnstone & Shanks, 2001; Newport & Aslin, 2004).

Although the experiment provided evidence of implicit knowledge of a two–step dependency, it is important to note that this may not actually be knowledge of subject–verb agreement. The participants may not have developed the abstract concepts of subject and verb. In fact, because every input sentence followed an identical structure, with the subject in the first position and the verb in the third, a sequential rule between words one and three could account for their performance without invoking any hierarchical structure or abstract linguistic concepts. The experiment did not allow us to distinguish between these options.

Although the non-detectors did become sensitive to the two–step dependency, this may have been the result of transfer. Subject–verb agreement exists in English (although there was no overt agreement in the translations provided, as they were in the past tense). The structure is also found in commonly taught languages such as French and German of which the participants had had some prior experience. Therefore it was not necessary for them to acquire the structure anew; they may simply have transferred it from previous knowledge of a different language. Unsurprisingly, when a structure is shared between languages, transfer from the first language often aids acquisition (Jarvis & Odlin, 2000; Odlin, 2003; Ringbom, 1992). Although Experiment 3 demonstrated that linguistic long–distance dependencies can be acquired implicitly, the role of transfer has to be clarified. Therefore Experiment 4 examined whether object–verb agreement (that is not found in the participants' L1 or previous L2s) is amenable to implicit learning when instantiated as a two–step dependency.

As reported in Section 5.1.2, the analysis of the ACS values in the timed test highlighted a confound between the independent variables of grammaticality and familiarity. Therefore knowledge of the latter may have been able to masquerade as knowledge of the former. However, if the participants had been responding on the basis of familiarity throughout and if therefore the grammaticality effect was merely the result of this confound, then a significant effect would also have been expected for familiarity. As there was none, the data supported the conclusion that the group had acquired the target structure but not the bigrams.

Once again none of the participants learnt the Persian bigrams implicitly, despite Experiment 1 suggesting that such patterns can be acquired and implicit knowledge of a different structure being acquired. The difficulty was therefore unlikely to be related to the learning task and more likely to be caused by the structure itself. Either language learners are less likely to learn lexically–based patterns than are participants in other types of experiment, or the familiar and unfamiliar sentences were not sufficiently different.

6.4 Chapter Summary

The source localisation task allowed participants to acquire implicit knowledge of a single two–step dependency when it was instantiated as subject–verb agreement in a novel language. However, as Experiment 2 but in contrast to Experiment 1, the frequent bigrams were not acquired. Finally, the rule detectors had explicit knowledge of the target rule and possibly also of the frequent bigrams.

CHAPTER 7

Experiment 4

The participants in Experiment 3 were able to acquire implicit knowledge of a two–step dependency based on subject–verb agreement. However they may have been transferring knowledge of the structure from their L1 or from previous L2s (Odlin, 2003). In Experiment 4 participants were trained on a dependency of the same length that could not have been learnt by transfer. The new target structure was object–verb agreement in a modified version of Basque. Object–verb agreement is less common cross–linguistically than subject–verb agreement and it is not found in the languages previously encountered by the participants.

7.1 Method

7.1.1 Participants

Two groups of eighteen participants (one trained and one untrained control group) were recruited in the same way and from the same population as was used for Experiments 2 and 3. None had taken part in the previous experiments. According to the responses on the language background questionnaire, the trained participants had encountered a median of two (interquartile range one to two) foreign languages prior to the experiment. The control participants on the other hand had learnt two (two to two). A two–tailed Mann–Whitney test with group (trained versus control) as the dependent variable and number of languages as the independent variable did not produce a significant effect, $U = 130.000$, ns. (Data were missing from one trained participant.) They also reported their current ability in their best foreign language on the same seven–point scale as previously. For the trained participants the median value was three (two to three), and for the control participants three (three to four). In each case the median equated to *I can make myself understood in certain limited practical situations*. Nevertheless a two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and ability as the dependent variable produced a significant effect, $U = 95.000$, $p = 0.043$. (Again, data were missing from one trained participant.) However the control participants reported higher ability levels than did the trained group, who could therefore be assumed to be at a relative disadvantage. Any findings of superior performance by the trained participants could not therefore be explained by greater levels of prior experience.

7.1.2 Materials

The stimuli employed in Experiment 4 were based on a modified version of Basque. The target structure was object–verb agreement for number in third person sentences. There were two versions of the stimuli (counterbalanced between participants), one with singular subject nouns and the other with plural subjects. Thus, the number value of the subject was held constant for each individual participant. Examples are shown in (20). Compare example *a* with *c* and *b* with *d*. In transitive sentences with a singular object noun, the auxiliary verb began with the prefix *du-*. When the object was plural it began *deetu-*.

(20)

a) Motily-e gison-a ikusi du-Ø
Boy-Ø girl-Ø seen has
Subj-sg Obj-sg Verb $\text{Aux}_{\text{ObjSg}}$ -$\text{Aux}_{\text{SubjSg}}$
"The boy saw the girl"

b) Motily-ek gison-a ikusi du-te
Boy-s girl-Ø seen have
Subj-pl Obj-sg Verb $\text{Aux}_{\text{ObjSg}}$ -$\text{Aux}_{\text{SubjPl}}$
"The boys saw the girl"

c) Motily-e gison-ak ikusi deetu-Ø
Boy-Ø girl-s seen has
Subj-sg Obj-pl Verb $\text{Aux}_{\text{ObjPl}}$ -$\text{Aux}_{\text{SubjSg}}$
"The boy saw the girls"

d) Motily-ek gison-ak ikusi deetu-te
Boy-s girl-s seen have
Subj-pl Obj-pl Verb $\text{Aux}_{\text{ObjPl}}$ -$\text{Aux}_{\text{SubjPl}}$
"The boys saw the girls"

The target items were constructed from three Basque nouns and three verbs selected from the vocabulary listed by Jansen (2002), so that all of the nouns

	Written Stems		Meaning	
	Original	**Modified**	**Original**	**Modified**
Nouns	Gizon	Gison	Man	Man
	Lagun	Lahun	Friend	Girl
	Mutil	Motily	Boy	Boy
Verbs	Agurtu	Ahurtu	Greet	Follow
	Ekarri	Ekarri	Bring	Find
	Ikusi	Ikusi	See	See

Table 7.1: Basque vocabulary. The spelling of the Basque stems was altered according to transcriptions produced by a native English speaker with no knowledge of Basque. The meanings were altered to improve the plausibility of the items.

contained two syllables and either five or six letters and phonemes. The verbs had three syllables and five or six letters and phonemes. Again it was not possible to select lexical items that met all of these requirements, and which also had appropriate meanings. Therefore some modifications were made.

There are substantial differences between the Basque and English spelling systems. To avoid participants having to learn novel grapheme–phoneme correspondences, the orthography was adapted. A naive native English speaker was asked to transcribe recordings of the Basque words spoken in isolation. The modified written forms used for the experiments were based on the resulting transcriptions.

The meanings of some of the words were also modified. The nouns had to be plausible both as the subject and as the object of a sentence. *Lagun* was altered from *friend* to *girl* because *friend* is usually used with a possessor. Verbs had to be monotransitive and be plausible with human subjects and objects. *Ekarri* was translated as *find* rather than *bring* as the latter is ditransitive, usually requiring an additional indirect object. Finally *follow* was used as the translation of *agurtu* in place of *greet*, which has an intransitive homophone in Scots meaning *cry*. Table 7.1 shows the original and modified forms.

In Basque, case and number are represented using nominal suffixes in a fusional system. The four that were used in this experiment are shown in Table 7.2. The singular subject affix was changed from *-ak* to *-e* to make the system agglutinative and therefore make the role of the target number agreement more transparent. In the resulting system case was indicated by the less salient change in the quality of the vowel, while the more salient presence or absence of the final consonant marked the number value of the noun. The word order was constant in the

	Original		Modified	
	Singular	**Plural**	**Singular**	**Plural**
Subject	-ak	-ek	-e	-ek
Object	-a	-ak	-a	-ak

Table 7.2: Basque nominal morphology. The *-ak* suffix used to mark singular subjects was altered to *-a* to create an agglutinative system whereby the vowel marked the case of the noun, and the final *-k* identified a plural.

		Subject			
		Singular		**Plural**	
		Original	**Modified**	**Original**	**Modified**
Object	**Singular**	Du: Du-Ø	N/A	Dute: Du-te	N/A
	Plural	Ditu: Ditu-Ø	N/A	Dituzte: Ditu-z-te	Ditute: Ditu-te

Table 7.3: Basque verbal morphology. The table uses the original spelling. As was the case for the lexical items, the spelling of *ditu* was changed to *deetu* according to the intuitions of a native English speaker. The *-z-* in *dituzte* was removed to create a regular agglutinative system whereby *ditu-* agreed with a plural object and *-te* with a plural subject.

experimental materials, so it was not necessary to process the case suffix in order to assign thematic roles.

The verbal inflections were carried on a sentence–final auxiliary rather than on the lexical verb. In the section of the paradigm that was employed, the auxiliary can be divided into two morphemes, the first representing the object agreement and the second the subject agreement (see Table 7.3). (Note that this segmentation is not valid for the entire paradigm in the natural language.) A singular subject was marked with a null suffix, and a plural subject by *-te*. *Du-* indicated a singular object, and *ditu-* a plural object. There is only one exception to the pattern, the obj_{plu}-$subj_{plu}$ form *dituzte* contains an extra *-z-* between the subject and object morphemes. For the sake of transparency, this *-z-* was omitted in the experimental materials leaving *ditute*.

The stimuli all had the same word order as was used in the examples in (20): subject — object — lexical verb — auxiliary. When creating an experimental item the subject noun was freely selected from the three used in the materials, while either of the remaining two could function as the object (thus avoiding repetitions). The object noun in Experiment 4 could be either singular or plural; the subject did not vary in number. The form of the auxiliary was entirely dependent on the number–value of the preceding nouns, and therefore did not permit any additional variation. Following these guidelines, it was possible to create thirty–six grammatical items for each experiment. In each case twenty–four were selected as the learning stimuli (the same number as in the biconditional grammar materials) with the additional restriction that there was one verb that could not follow each object noun. Regardless of whether it was singular or plural, *motily* was not followed by *ekarri*, *gison* was not followed by *ahurtu*, and *lahun* was not followed by *ikusi*. The learning items are listed in Appendix F.

The forty–eight judgement test items can be seen in Appendix F (as with the biconditional grammar materials the same stimuli were used for the timed and for the untimed test). As previously, grammaticality and familiarity were both manipulated. One quarter of the items were G-F, one quarter G-NF, one quarter NG-F, and the final quarter NG-NF. In the ungrammatical stimuli there was a number mismatch between the auxiliary and the object. The unfamiliar sentences contained a novel object–lexical verb bigram. The novel bigram had not been used in the learning items, and nor had any other one using the same lexical items and differing only in inflection. Thus, if *lahuna ekarri* was a novel bigram, then neither it nor *lahunak ekarri* were used in the learning phase items. This was equivalent to a lexically novel bigram in the Persian materials. Again, familiarity was a one–step dependency between adjacent words.

In the timed test, the G-F items had a mean z-transformed ACS value of 1.43, the G-NF items of -0.92, the NG-F items of 0.41 and the -0.92. The corresponding figures for the untimed test were G-F 1.71, G-NF -0.57, NG-F -0.57, NG-NF -0.57. These were calculated using the combined frequencies of each bigram and trigram in the learning phase and in the timed test, following from the assumption that learning continues during the test. Finally, the values in the post–experiment analyses, based on the exposure in the learning phase and both judgement tests, were G-F 0.19, G-NF 0.74, NG-F -1.67, and NG-NF 0.74. No inferential analyses were carried out because the variance within each group was zero throughout. There was a contrast between the familiar and the unfamiliar stimuli in the timed test, which was reduced to a contrast between the G-F items and the remaining three types in the untimed test. Finally, by the post–experiment calculations the contrast between the familiar and unfamiliar items was in the opposite direction to that predicted.

The correction test used the twelve items shown in Appendix F, of which six were fillers with word order violations. The six experimental items were all NG-NF. In each there was a violation of the target agreement rule, and an object–lexical verb bigram that had not been employed in the learning items. One of the four suggested corrections concerned the number morpheme on either the relevant noun or the auxiliary. Thus, it improved the sentence from NG-NF to G-NF. A second correction altered the lexical item in either the object or the verb slot, creating a familiar bigram. If this option was selected, the item became NG-F. This allowed any familiarity detectors to be identified. A third suggested correction concerned the word order, and the fourth the lexical item in the subject slot.

As in the Persian case, the spoken stimuli were constructed by concatenating tokens of the individual words without any further processing. A male speaker of Basque recorded three tokens of each word in isolation, and the best ones were selected to make the spoken sentences. The gaps between words were cut so that the resulting stimuli lasted exactly three seconds. Nevertheless brief pauses remained between each word.

7.1.3 Procedure

The participants in Experiment 4 were trained using the source localisation task, as in Experiment 3. However the target items in the source localisation task were presented for 5700 ms rather than for 4700 ms (the exposure length for the practice trials with the individual words was held at 2000 ms). The new exposure time was based on a pilot experiment, in which participants were instructed to press the space bar immediately after reading an item in each language. The resulting reading time measures indicated that the Persian items (which contained three words) took a mean of 2493 ms to read (standard error 336). The Basque items (which contained four words) had a mean reading time of 3032 ms (297). Two t-tests with language (Persian versus Basque) as the dependent variable and reading time as the dependent variable confirmed that there was a significant difference between the two, $t_1(4) = 6.370$, $p = 0.003$, $\eta_p^2 = 0.910$, $t_2(62) = 4.737$, $p < 0.001$, $\eta_2^2 = 0.266$. The Basque items took 21% more time to read than did the Persian ones, and therefore the exposure time was also lengthened by 21%. In all other respects the procedure was the same for the two experiments.

7.1.4 Analysis

The correction test allowed the option to respond on the basis of familiarity. Therefore it was used to divide the trained participants into three groups (as in Experiment 1): rule detectors, familiarity detectors, and non-detectors. The first group reached the criterion of responding to four or more of the six experimental

	Control	Non-Detectors	Rule Detectors	Familiarity Detectors
Grammaticality	2.28 (0.28)	2.38 (0.22)	5.11 (0.25)	2
Familiarity	1.39 (0.23)	1.63 (0.28)	0.33 (0.22)	4

Table 7.4: Experiment 4 correction test performance. The table shows the mean (and standard error) of the number of items corrected on the basis of grammaticality (from NG-NF to G-NF) and on the basis of familiarity (from NG-NF to NG-F) for each group of participants. There was only one familiarity detector, and therefore no variance in their performance.

items on the basis of grammaticality. The second group reached the criterion by responding on the basis of familiarity. The final group did not reach the criterion in either way. As previously the criterion was set so that no more than 5% of participants would be expected to achieve it by chance. Each trained group's judgement test performance was then contrasted with that of the control participants in the same way as previously, to investigate whether they had acquired implicit or explicit knowledge of either the target structure or the frequent bigrams.

7.2 Results

7.2.1 Correction Test Results

Following their correction test performance. Ten of the trained participants were classified as rule detectors, one was a familiarity detector, and the remaining seven were non-detectors. The number of correction test items answered according to grammaticality and familiarity is shown in Table 7.4. As there was only a single familiarity detector, their data were not analysed further.

7.2.2 Learning Phase Accuracy

As in Experiment 3, the trained participants in Experiment 4 carried out the source localisation task during the learning phase. Their overall mean accuracy was 66.55% (standard deviation 1.64). Performance in the learning phase was also calculated separately for each group of trained participants. The non-detectors had a mean accuracy in the source localisation task of 67.56% (9.31). The corresponding figures for the rule detectors were 65.63% (5.58). A two–tailed t-test with group (non-detector versus rule detector) as the independent variable and learning phase accuracy as the dependent variable did not produce

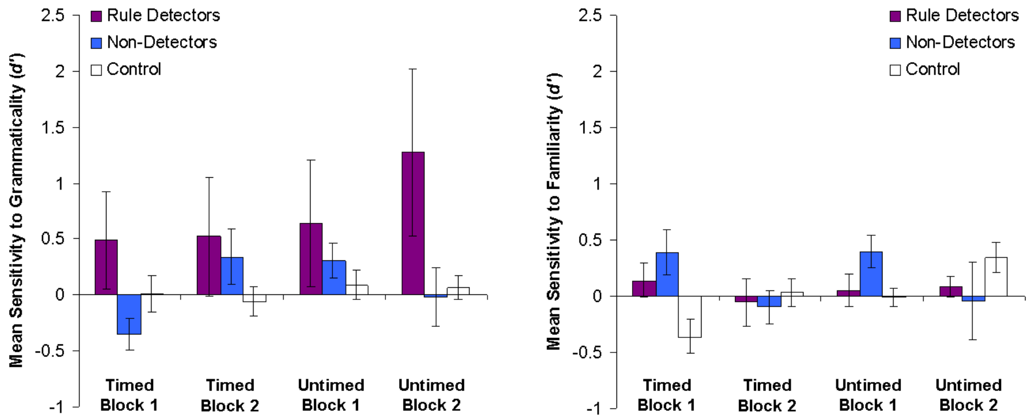


Figure 7.1: Experiment 4 results. The figure shows each group of participant's mean sensitivity to the dependent variables as measured in d' units. Error bars show the standard error.

a significant effect, $t < 1$. Therefore the two groups of trained participants performed equivalently in the source localisation task.

7.2.3 *Judgement Test Results*

Figure 7.1 shows each group's sensitivity to the two independent variables (with the exception of the familiarity detector). As previously, ANOVAs were used to compare each group of trained participants' performance against that of the controls.

The non-detectors had acquired the two–step dependency in Experiment 3, and so they were expected to do the same in Experiment 4. However an ANOVA with condition (non-detector versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality in the timed test as the dependent variable showed that there was neither a significant main effect of condition, $F < 1$, nor one of block, $F(1, 23) = 2.438$, ns, $\eta_p^2 = 0.096$. However there was a marginal interaction, $F(1, 23) = 3.561$, $p = 0.072$, $\eta_p^2 = 0.134$. Figure 7.1 shows that the non-detectors were less sensitive to grammaticality in block one of the timed test than were the control participants, removing the need to test the data statistically and also suggesting that an increase in the number of participants would be unlikely to lead to a significant effect. A t-test for block two did not produce an effect of condition either, $t(23) = 1.537$, ns, $\eta_p^2 = 0.093$, indicating that the non-detectors were no more sensitive to grammaticality than were the control participants. The corresponding ANOVA for the untimed test did not reveal any significant effects, all Fs < 1. In summary, there is no evidence that the group had developed any knowledge of the target two–step object–verb agreement, implicit or otherwise. This is in contrast to the non-detectors in Experiment 3 who acquired subject–verb agreement of the same length. If the change in structure were the cause of this pattern of results, it would suggest that the participants had acquired the abstract linguistic structures.

As the familiarity manipulation had been strengthened relative to the Persian materials and the reliance on inflected novel bigrams removed, the non-detectors were predicted to acquire the frequent bigrams. An ANOVA with sensitivity to familiarity in the timed test as the dependent variable produced a marginal main effect of condition, $F(1, 23) = 3.197$, $p = 0.087$, $\eta_p^2 = 0.122$, but not one of block, $F < 1$. However there was also a significant interaction, $F(1, 23) = 5.839$, $p = 0.024$, $\eta_p^2 = 0.202$. Dividing the test into two blocks for the planned comparisons highlighted a significant effect of condition in block one, $t(23) = 2.694$, $p = 0.007$, $\eta_p^2 = 0.240$, but not in block two, $t < 1$. Neither main effect was significant in the untimed–test ANOVA, all Fs < 1, but the interaction remained so, $F(1, 23) = 5.529$, $p = 0.028$, $\eta_p^2 = 0.194$. Further planned comparisons confirmed that the effect was again significant in block one, $t(23) = 2.585$, $p = 0.009$, $\eta_p^2 = 0.225$, but not in block two $t(23) = -1.248$, ns, $\eta_p^2 = 0.063$.

Although the same set of novel bigrams was used in both tests, each used a different modality. Judgements of spoken stimuli became less accurate following exposure to the spoken novel bigrams in the timed test during which the participants presumably acquired those bigrams. On the other hand performance with the written stimuli appeared to remain unaffected until the written novel bigrams were encountered and acquired in the untimed test. The pattern of results suggests that implicit knowledge is represented in a modality–specific form, at least to some extent. If this assumption is valid, the familiarity effect met the three criteria for implicit learning. The non-detectors did not complete the correction test according to familiarity, and they were sensitive to familiarity in the timed test but not in block two of either judgement test.

The rule detectors were defined as performing at above chance levels in the correction test, which was assumed to require explicit knowledge. They were not expected to be able to access this knowledge under time pressure. The ANOVA conducted with sensitivity to grammaticality in the timed test as the dependent variable did not produce any significant effects, condition $F(1, 26) = 2.096$, ns, $\eta_p^2 = 0.075$, all other Fs ≤ 1. The parallel ANOVA for the untimed test revealed a marginal main effect of condition, $F(1, 26) = 3.432$, $p = 0.075$, $\eta_p^2 = 0.117$, indicating a contrast between the rule detectors and the controls, but neither a main effect of block, $F(1, 26) = 2.336$, ns, $\eta_p^2 = 0.082$, nor a condition x block interaction, $F(1, 26) = 2.336$, ns, $\eta_p^2 = 0.093$. Thus as predicted, the only statistical evidence from the judgement tests that the rule detectors had learnt the target structure was in the absence of time constraints, but the effect did not reach significance.

A final pair of ANOVAs assessed whether the rule detectors had acquired sensitivity to familiarity. The timed test version did not reveal any significant effects, condition $F(1, 26) = 1.989$, ns, $\eta_p^2 = 0.071$, block $F < 1$, condition x block $F(1, 26) = 2.634$, ns, $\eta_p^2 = 0.092$. The same was true of the ANOVA for the untimed test, condition $F < 1$, block $F(1, 26) = 1.887$, ns, $\eta_p^2 = 0.068$, condition x block $F(1, 26) = 1.291$, ns, $\eta_p^2 = 0.047$. Thus, there is no evidence that the rule detectors had learnt the frequent bigrams.

7.2.4 Response Biases

The non-detectors in Experiment 4 were sensitive to familiarity in the first block of each judgement test but not in the second. This finding was consistent with the interpretation that they continued to acquire implicit knowledge during the test, thereby integrating the novel bigrams into their grammar. As has been described above, unless the threshold above which items were classified as grammatical was also changed, this process would lead to an increase in the proportion of items that were accepted. Therefore the values of the response bias (*c*) were expected to increase during the experiment.

Figure 7.2 shows the progression in the values of *c*, suggesting that there was indeed a gradual increase. A one–way ANOVA with block (timed one, timed two, untimed one, untimed two) as the independent variable and *c* as the dependent variable produced a significant effect $F(3, 18) = 2.680$, $p = 0.039$, $\eta_p^2 = 0.309$. As predicted but in contrast to Experiments 1 and 3, exposure to the incorrect judgement test stimuli did appear to broaden the non-detectors' grammar. Nevertheless by the end of the experiment the biases were still negative — the participants were more likely to reject than to accept the experimental items regardless of their status.

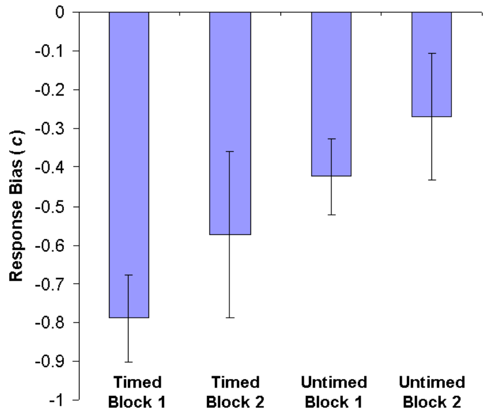


Figure 7.2: Mean response biases in Experiment 4. The non-detectors were sensitive to familiarity. Therefore the figure shows the response biases irrespective of the familiarity of an item, with responses of *correct* to familiar items considered hits and responses of *correct* to unfamiliar items labelled false alarms.

7.3 Discussion

In contrast to Experiment 3, the non-detectors did not acquire the target two–step dependency. The earlier target structure was subject–verb agreement, which could have been transferred from the L1 or previous L2s. The current one (object–verb agreement) was novel. Thus one interpretation is that linguistic long–distance dependencies can only be acquired through transfer, which would imply that the participants do acquire the underlying structures. Transfer is unlikely to have been the only means by which explicit knowledge was acquired however, because a larger percentage of the trained participants became rule detectors (56%) than in Experiment 3 (36%), where the target structure was subject–verb agreement. They were more likely to acquire explicit knowledge of the non-transferable than of the transferable structure, and were therefore unlikely to have been relying largely on transfer. Nevertheless, the same may not necessarily be the case for the acquisition of implicit knowledge, and an alternative explanation may also be offered.

Implicit knowledge of the frequent bigrams was not learnt in Experiment 3 for independent reasons, so the target rule was the shortest (and possibly therefore the simplest) regularity in the data. However this was not the case in the Basque data set, where the familiarity pattern was acquired. Learning the simpler regularity may have prevented the participants from acquiring the longer rule, at least within the experimental time frame. Supporting this interpretation, Cleeremans and McClelland (1991) and Schvaneveldt and Gomez (1998) reported that longer contingencies were acquired more slowly than smaller ones. Section 10.2.4 reports a simulation with a longer exposure period to test this possibility.

One of the explanations suggested to account for the non-detectors' lack of sensitivity to familiarity in Experiment 3 was that language learners are restricted to acquiring word–class based patterns rather than regularities based on lexical items. However the non-detectors in Experiment 4 acquired implicit knowledge of the frequent bigrams and this pattern was based on individual lexical items rather than on categories. Thus, the current findings are not consistent with the word–class account. It is more likely either that the frequent and novel bigrams were too difficult to distinguish in the Persian materials, or that the ACS contrast between the categories was insufficient to allow familiar and unfamiliar strings to be identified. The latter option will be tested by simulation in Section 10.2.4.

The non-detectors' sensitivity to familiarity was above that of the controls in block one but not in block two of the timed test, thereby conforming with the criteria for implicit knowledge. However contrary to the expectations the same also occurred in the untimed test, where accuracy improved in block one but returned to chance in block two. This pattern of results could be explained if the participants acquired the spoken novel bigrams in the timed test and integrated them into their grammar, but did not acquire the written novel bigrams until the untimed test. Thus, this explanation requires their knowledge of the auditory target language forms to be separate from their knowledge of the written forms. This finding is consistent with Schacter and Graf (1989) who found that a change in modality eliminated implicit memory for new associations between unrelated words, suggesting that representations may be modality-specific early in the learning process. Nevertheless, a thorough investigation of the separability of spoken and written representations in SLA is outside the remit of the current thesis.

7.4 Chapter Summary

Without the option to transfer from the L1 implicit knowledge of the target two–step dependency was not acquired, thus implying that learners may not be able to acquire new linguistic long–distance dependencies. On the other hand, the non-detectors acquired the frequent bigrams, thus confirming that implicit knowledge of lexically–based patterns can be acquired from language input. It is therefore possible that the non-detectors' success with the frequent bigrams prevented them from also acquiring the longer dependency during the short exposure period.

CHAPTER 8

Experiment 5

Experiment 3 demonstrated that implicit knowledge of a two–step dependency can be acquired, while Experiment 4 found evidence of implicit knowledge of one–step regularities in a new foreign language. Currently there is no evidence on the upper limit of dependency that can be acquired from language input. The current experiment examines the case of three–step dependencies, again using linguistic input. This is outside the limit proposed for other experimental paradigms (Cleeremans & McClelland, 1991; Gómez, 1997), and so success would not be predicted from a domain–general process.

The target structure in Experiment 5 was subject–verb agreement, as in Experiment 3. Therefore participants could perform accurately by transferring implicit or explicit knowledge of the rule from their L1 (English) or from previous L2s. They did not need to learn the structure anew. If transfer is a sufficient condition, then successful implicit knowledge should be identified in the current experiment.

8.1 Method

8.1.1 Participants

Eighteen participants were recruited from the same population as in Experiments 2, 3 and 4. An equal number of control participants were drawn from the same population. None had taken part in the previous experiments.

The language background questionnaire was examined to judge whether the trained and control participants had similar levels of prior exposure to foreign languages. The trained participants had encountered a median of two foreign languages (interquartile range one to two). The control participants had encountered two (two to three). A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and number of languages as the dependent variable produced a marginal effect, $U = 102.000$, $p = 0.059$. (Data were missing for one trained participant.) The trained participants had encountered fewer foreign languages on average than had the controls, and could therefore be assumed to be at a relative disadvantage.

The trained participants had a median ability in their best foreign language of three (two to three), translating as *I can make myself understood in certain limited practical situations* on the same seven–point scale used previously. The corresponding figure for the control participants was also three (three to four). A two–tailed Mann–Whitney test with group (trained versus control) as the independent variable and ability as the dependent variable did not result in a significant effect, $U = 122.500$, ns. (Data were missing for one trained participant.) Therefore the two groups of participants had reached comparable ability levels in their previous foreign languages.

8.1.2 Materials

The stimuli in Experiment 5 were based on the same modified Basque materials used for Experiment 4. However, instead of object–verb agreement the new target structure was subject–verb agreement. Half the participants saw only singular objects, and the remainder saw only plural objects. Thus, for each individual participant the number value of the object was invariant.

The example sentences in (21) are repeated from (20). The effects of the target subject–verb agreement for number can be seen by comparing example *a* with example *b* and *c* with *d*. In summary, plural subjects required an overt agreement suffix on the auxiliary verb (*-te*), but singular subjects did not. Using the same procedure as in Experiment 4, twenty–four learning items were created.

(21)	a)	Motily-e	gison-a	ikusi	du-Ø
		Boy-Ø	girl-Ø	seen	has
		Subj-sg	Obj-sg	Verb	$\text{Aux}_{\text{ObjSg}}$ -$\text{Aux}_{\text{SubjSg}}$
		"The boy saw the girl"			

b)	Motily-ek	gison-a	ikusi	du-te
	Boy-s	girl-Ø	seen	have
	Subj-pl	Obj-sg	Verb	Aux_{ObjSg} -Aux_{SubjPl}
	"The boys saw the girl"			
c)	Motily-e	gison-ak	ikusi	deetu-Ø
	Boy-Ø	girl-s	seen	has
	Subj-sg	Obj-pl	Verb	Aux_{ObjPl} -Aux_{SubjSg}
	"The boy saw the girls"			
d)	Motily-ek	gison-ak	ikusi	deetu-te
	Boy-s	girl-s	seen	have
	Subj-pl	Obj-pl	Verb	Aux_{ObjPl} -Aux_{SubjPl}
	"The boys saw the girls"			

The forty–eight judgement test items can be seen in Appendix F. As in Experiments 1 and 4 the same stimuli were used for the timed and for the untimed test, and one quarter of the items were G-F, one quarter G-NF, one quarter NG-F, and the final quarter NG-NF. In the ungrammatical stimuli there was a number mismatch between the auxiliary and the subject. The unfamiliar stimuli contained a novel bigram that consisted of the object and the lexical verb. This novel bigram had not been used in any of the learning stimuli and nor had any other bigram consisting of the same two lexical items and differing solely in inflection. Thus, if *lahuna ekarri* was a novel bigram, then neither it nor *lahunak ekarri* had been used in the learning phase items.

In the timed judgement test the mean z-transformed ACS value for the G-F items was 0.99, for the G-NF items -0.99, for the NG-F items 0.99, and for the NG-NF items -0.99. Within each type, every item had exactly the same ACS so the variance was zero. Thus, it was not necessary to calculate inferential statistics in order to claim that the familiar and unfamiliar items differed significantly in this regard, but that the grammatical and ungrammatical items did not. The chunk strengths were identical for the untimed test based on exposure during the learning phase and the timed test, so the same conclusions can be drawn. The direction of the effect had reversed by the post–experiment calculation based on the learning phase and both judgement tests, with the value for both the G-F items -0.99, for the G-NF items 0.99, for the NG-F items -0.99, and for the NG-NF items 0.99.

The correction test used the twelve items shown in Appendix F, of which six were fillers with word order violations. The six experimental items were all NG-NF, as in Experiment 4. One of the four suggested corrections concerned the number morphemes, improving the sentence from NG-NF to G-NF. A second correction altered the lexical item in either the object or the verb slot, creating a familiar bigram to make the item NG-F. One foil concerned the word order, and the second the lexical item in the subject slot.

Finally, the spoken versions of the stimuli were created in the same way and from the same recordings of the individual words as in Experiment 4.

8.1.3 Procedure

The procedure was identical to that in Experiment 4 in all respects.

8.1.4 Analysis

Following the same procedure as in Experiment 4, the correction test was used to divide the trained participants into three groups: rule detectors, familiarity detectors, and non-detectors. As there were six experimental items, the criterion for classification as a rule– or familiarity detector was again set at four or more items correct. Each group's judgement test performance was then compared to that of the untrained control participants in the same way as previously.

8.2 Results

8.2.1 Correction Test Results

Nine of the eighteen trained participants in Experiment 5 were classified as rule detectors according to the criterion, one was classified as a familiarity detector, and the remaining eight were non-detectors. Table 8.1 shows the correction test performance of each trained group, and also that of the control participants. As there was only one familiarity detector their performance was not analysed further.

8.2.2 Learning Phase Accuracy

The trained participants in Experiment 5 carried out the same source localisation task as in Experiments 3 and 4. Their mean learning phase accuracy was 65.74% (standard deviation 2.39). The source localisation responses of the two remaining groups of trained participants in Experiment 5 (the non-detectors and

	Control	Non-Detectors	Rule Detectors	Familiarity Detectors
Grammaticality	3.56 (0.27)	1.88 (0.15)	5.33 (0.29)	2
Familiarity	0.83 (0.20)	1.50 (0.22)	0.56 (0.29)	4

Table 8.1: Experiment 5 correction test performance. The table shows the mean (and standard error) of the number of items corrected on the basis of grammaticality (from NG-NF to G-NF) and on the basis of familiarity (from NG-NF to NG-F) for each group of participants. There was only one familiarity detector and therefore no variance in their performance.

rule detectors) were also contrasted. The former had a mean accuracy level of 64.84% (11.16), and the latter of 65.28% (9.66). Unsurprisingly, a two–tailed t-test with group (non-detector versus rule detector) as the independent variable and learning phase accuracy as the dependent variable did not reveal a significant effect, $t < 1$. The two groups therefore performed at the same level in the source localisation task.

8.2.3 Judgement Test Results

Figure 8.1 shows each group's sensitivity to grammaticality and familiarity, with the exception of the single familiarity detector. It indicates that the non-detectors were less sensitive to grammaticality than were the control participants in both blocks of the timed test, making statistical testing unnecessary. An ANOVA with condition (non-detector versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable confirmed that there were no significant effects in the untimed test either, all Fs < 1. The group did not acquire the target structure.

As in Experiment 4, the familiarity manipulation was based purely on the lexical items, not on the inflected forms as was often the case in Experiments 2 and 3. Therefore the non-detectors were expected to acquire the frequent bigrams. However an ANOVA with sensitivity to familiarity in the timed test as the dependent variable did not demonstrate any significant effects, all Fs < 1. The same was true for the untimed test ANOVA, condition $F(1,24) = 2.179$, ns, $\eta_p^2 = 0.083$, block $F(1,24) = 1.507$, ns, $\eta_p^2 = 0.059$, condition x block $F < 1$. Surprisingly and in contrast to Experiment 4, the non-detectors did not acquire the frequent bigrams.

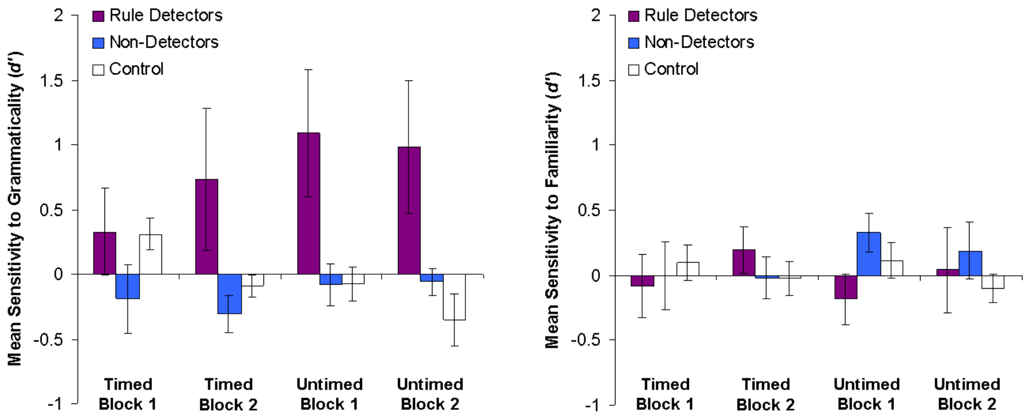


Figure 8.1: Experiment 5 results. The figure shows each group of participant's mean sensitivity to the dependent variables as measured in d' units. Error bars show the standard error.

The rule detectors demonstrated that they had learnt the target structure in the correction test, but they were not expected to be able to access that knowledge in the timed test. An ANOVA with condition (rule detectors versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable did not produce a significant effect of condition, $F(1, 25) = 3.965$, ns, $\eta_p^2 = 0.137$, nor one of block, $F < 1$, but there was a significant interaction, $F(1, 25) = 7.805$, $p = 0.010$, $\eta_p^2 = 0.238$. Planned comparisons did not produce a significant effect of condition in block one, $t < 1$, or in block two, $t(25) = 1.475$, ns, $\eta_p^2 = 0.144$. They confirmed that the rule detector group was no more sensitive to the target structure than were the control participants. The untimed test reflected a different pattern. The ANOVA revealed a significant effect of condition, $F(1, 25) = 10.339$, $p = 0.004$, $\eta_p^2 = 0.293$, but neither a main effect of block, $F(1, 25) = 1.258$, ns, $\eta_p^2 = 0.048$, nor an interaction, $F < 1$. Thus, the rule detectors performed above chance in the correction test, they were not sensitive to grammaticality in the timed test, and their sensitivity to grammaticality was maintained throughout the untimed test. Thus, their performance met the criteria to be classified as the result of explicit knowledge.

Like the non-detectors, the rule detectors did not respond to the correction test items on the basis of familiarity. As the former group did not demonstrate any knowledge of the frequent bigrams in the judgement tests, the same was expected to be true for the latter. Consistent with this, an ANOVA with sensitivity to familiarity in the timed test as the dependent variable did not produce significant main effects of either condition or block, all Fs < 1, nor a condition x block interaction, $F(1, 25) = 1.739$, ns, $\eta_p^2 = 0.065$. The same was true for a parallel

ANOVA using the untimed test performance, condition x block $F(1, 25) = 2.118$, ns, $\eta_p^2 = 0.078$, all other Fs < 1. Neither the rule detectors nor the non-detectors acquired the frequent bigrams.

8.2.4 Attraction Errors

In Experiments 2 and 3 where attraction error analyses were also conducted, each participant saw both singular and plural local nouns. This was not the case in Experiment 5. In the current Basque materials the object (the local noun) also triggers agreement on the auxiliary. To hide the effect of this dependency in the materials for Experiment 5, each participant received either singular or plural objects, but not both. As the number value of the object was constant for each participant in Experiment 5, attraction errors may have been less likely to occur.

Nevertheless, sensitivity to grammaticality was calculated separately for the number match and mismatch items in the same way as for Experiments 2 and 3 in Sections 5.2.5 and 6.2.6. The results are shown in Figure 8.2. Unlike in the previous experiments, the figure demonstrates that the control participants differed substantially and non-randomly from chance, and therefore that they did not provide a good baseline for statistical analysis.

Figure 8.2 shows that the non-detectors' mean sensitivity to grammaticality was almost always within one standard error of chance (with the exception of the mismatch stimuli in block two of the timed test). If they had learnt the target structure but were prevented from demonstrating their knowledge because of attraction errors, then they should have been above chance in the match items, at least in block one of the timed test. According to Figure 8.2, this was not the case.

The rule detectors acquired the target structure, but their performance was not significantly above chance in the timed test. If they made a large number of attraction errors when under time constraints, it may have been sufficient to mask otherwise accurate performance with the match items. However Figure 8.2 indicates that their mean performance with each type of stimulus was never more than one standard error apart. Attraction errors are therefore unlikely to have substantially affected the results.

Overall, there is little evidence that the trained participants in Experiment 5 made sufficient attraction errors to obscure any knowledge of the target structure. Therefore the previous conclusions from Section 8.2.3 remain unchallenged. The

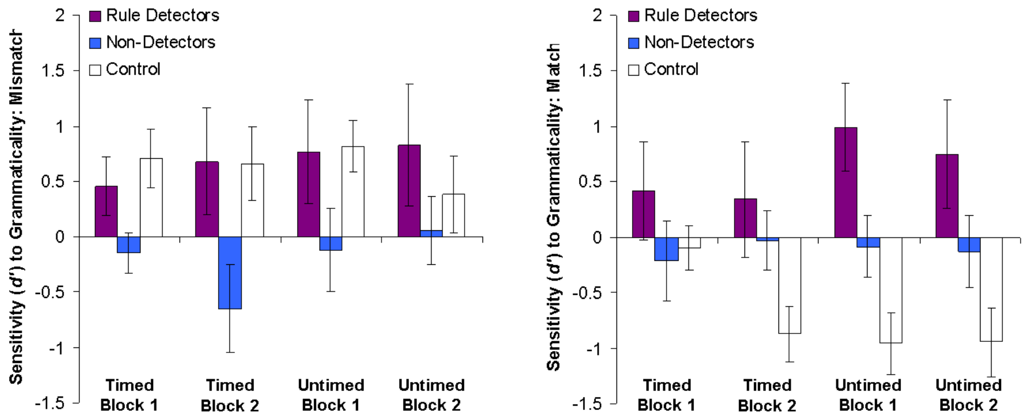


Figure 8.2: Mean sensitivity to grammaticality in Experiment 5 in the number–match and number–mismatch items. The error bars show the standard error.

non-detectors did not appear to acquire the target structure, while the rule detectors learnt subject–verb agreement but were not able to use their knowledge reliably in the timed test.

8.3 Discussion

Unlike in Experiment 3, there was no evidence that the participants acquired the target structure. Lengthening the dependency from two to three steps appeared to prevent the acquisition of implicit knowledge, despite the availability of transfer, as the target structure was subject–verb agreement in both cases. The argument is strengthened by the participants' insensitivity to the frequent bigrams. In contrast to Experiment 4, their failure to learn the target structure cannot be explained by them learning the shorter regularity. Thus, two–step patterns appear to be the longest that can be known implicitly even in adult second language acquisition. Note that, as transfer would have led to significant grammaticality effects, Experiment 5 also suggests that the availability of transfer is not a sufficient condition for implicit knowledge to be acquired.

A two–step length limit is compatible with the findings from domain general research in the AGL paradigm. Gómez (1997) found that trigrams can be acquired implicitly, while four–item chunks require explicit learning. (These patterns contain the same number of elements as two– and three–step dependencies respectively.) The four–step dependencies produced by a biconditional grammar were not acquired in Experiment 1 and contingencies of the same length have not been acquired in previous research either (Cleeremans & McClelland, 1991;

Johnstone & Shanks, 2001; Mathews et al., 1989). Thus, the current results are consistent with findings in other experimental paradigms.

The participants in Experiment 4 became sensitive to the frequent bigrams, thereby proving that lexical patterns can be acquired by language learners. The same was not the case here despite the fact that both experiments used similar Basque materials. There was a comparable ACS contrast between the familiar and unfamiliar items in each experiment and neither set required participants to pay attention to the agreement morphology in order to judge whether a sentence contained a novel bigram. However there was a difference between the two data sets in the overlap between the familiarity and grammaticality manipulations. All the novel bigrams in the Basque data consisted of the object noun and the lexical verb. In Experiment 4, which tested object–verb agreement, one of those words was also involved in the target structure. In Experiment 5, which focused on subject–verb agreement, neither of the words involved in the novel bigrams underwent any morphological variation. It may be that the participants paid less attention to the morphologically invariant parts of the sentence and that this lack of attention prevented them either from acquiring or from later demonstrating knowledge of the frequent bigrams. The role of attention in the development of implicit knowledge is discussed in Section 11.2.3.

8.4 Chapter Summary

The target three–step subject–verb agreement rule was not acquired by the non-detectors, suggesting that implicit knowledge of contingencies with more than two steps is not acquired, and demonstrating that transfer is not a sufficient condition for the phenomenon to take place. Surprisingly there was no effect of the frequent bigrams, with participants not distinguishing between familiar and unfamiliar items. A tentative explanation was offered based on attention, which will be developed in more detail in Section 11.2.3.

Chapter 9

Secondary Analyses

The analyses reported in the previous chapters considered each of the experiments individually. This chapter reports three additional analyses that included data from multiple experiments. First, memorisation performance may have differed between materials or learning tasks and thereby affected overall learning outcomes. Therefore Section 9.1 contains a comparison of accuracy levels in the learning phases of the five experiments. Second, although the participants in the language learning experiments (Experiments 2 — 5) had minimal foreign language knowledge, there was nevertheless some interparticipant variation. Whether superior experimental performance was associated with greater prior experience is assessed in Section 9.2. Third, participants trained in explicit conditions have previously been shown to respond more slowly than those in implicit conditions in judgement tests (Bitan & Karni, 2004). Section 9.3 discusses whether there was a similar difference between the rule detectors and the non-detectors in the current experiments.

9.1 Comparing Learning Phase Accuracy

Conceivably, the different results in the five experiments could have resulted either from contrasting abilities to acquire the abstract structures or simply from different memorisation capabilities. This section reports an analysis of learning phase performance to test this possibility. The trained participants in Experiments 1 and 2 performed the reproduction task during the learning phase, retyping each item from memory once it was no longer present on the monitor. Those in Experiments 3 — 5 carried out the source localisation task, matching items to speakers. Figure 9.1 shows the mean learning phase performance for each experiment, indicating

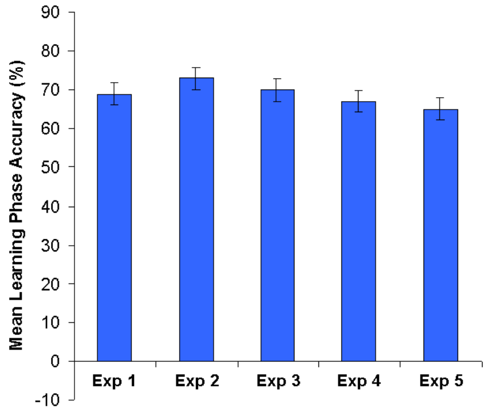


Figure 9.1: Mean learning phase accuracy across experiments. The error bars show the standard error. Experiments 1 and 2 used the reproduction task and Experiments 3 — 5 the source localisation task. Reproduction task responses had to be 100% accurate in order to be classified as correct.

that performance was equivalent regardless of the learning task. An ANOVA with experiment as the independent variable and learning phase accuracy as the dependent variable did not produce a significant effect, $F(4, 119) = 1.353$, ns, $\eta_p^2 = 0.044$, confirming that there were no significant differences between experiments. Therefore the contrast in learning outcomes reported in the preceding chapters was more likely to have resulted from differences in the ability to abstract the target structures.

9.2 Language Background

The amount of previous experience participants had learning foreign languages may have influenced their performance in the experiment. As described above, a questionnaire administered at the end of the procedure asked which languages a participant had previously learnt, and it also asked them to rate their current ability level in their best foreign language on a seven point scale (see Section 4.1.1). Sections 4.1.1, 5.1.1, 6.1.1, 7.1.1 and 8.1.1 reported comparisons between the trained and control participants' questionnaire responses. In contrast, this section only considers the trained participants and asks whether the rule detectors had more prior experience than did the non-detectors. Only the participants who were exposed to linguistic stimuli were included (i.e. Experiment 1 was excluded from the analyses). Similarly, because there were so few familiarity detectors those groups were also excluded. To maximise power, the data were collapsed across Experiments 2 — 5.

	Rule Detectors	Non-Detectors
Number of Languages	2 (1 — 2)	2 (1 — 2)
Best Ability	3 (2.4 — 3.5)	3 (2 — 3)

Table 9.1: Rule detectors' and non-detectors' median foreign language experience prior to the experiment (with interquartile ranges in parentheses).

Table 9.1 summarises the two language background variables for the rule detectors and non-detectors separately. A one–tailed Mann–Whitney test demonstrated that the rule detectors had learnt significantly more foreign languages than the non-detectors $U = 1102.500$, $p = 0.037$. Similarly, a one–tailed Mann–Whitney test indicated that they had reached significantly higher ability levels, $U = 1093.000$, $p = 0.044$. Overall therefore, the trained participants who became rule detectors had more experience with foreign languages than those who became non-detectors.

The previous analyses asked whether there was a significant difference between the non-detector and rule detector groups in terms of previous language learning experience. However the participants' foreign language backgrounds may also have led to differences in performance *within* each of the trained groups. That is, rule detectors who had greater previous experience may have outperformed rule detectors who had less. This was tested by means of separate multiple regression analyses for each group in each test. Sensitivity to each of the independent variables was the dependent variable and the two language background variables were the predictors[1]. As we were not aware of theoretical reasons for entering one of the predictor variables before the other, a backwards stepwise method was employed. That is, both predictors were initially entered into the analysis before those that did not significantly contribute to the success of the model were removed one–by–one in the following steps (Field, 2005).

None of the regression analyses for the non-detectors revealed a significant effect. Thus, neither language background variable explained a significant portion of the variance in their sensitivity either to grammaticality or to familiarity in either test. The same was true for the rule detectors in the timed test, and for their sensitivity to familiarity in the untimed one. However, best ability was shown to be a significant predictor of the rule detectors' sensitivity to grammaticality in the untimed test. (See Table 9.2 for further details.) Thus, previous language learning

[1]A Spearman's ρ test highlighted a significant correlation between the two predictor variables, $\rho = 0.409$, $p < 0.001$. Nevertheless the correlation coefficient was below the 0.8 threshold that Field (2005) recommended as the maximum for multiple regression analyses.

	B	*SE B*	β
Step 1			
Constant	-0.930	0.785	
No. Langs.	0.248	0.291	0.147
Ability	0.386	0.290	0.229
Step 2			
Constant	-0.839	0.775	
Best Ability	0.515	0.247	0.306*

Table 9.2: Backward multiple regression. The analyses assessed the rule detectors' sensitivity to grammaticality in the untimed test relative to their language background. $R^2 = 0.109$ for Step 1; $\Delta R^2 = -0.016$ for Step 2, all ns. * $p = 0.043$.

experience only appeared to affect performance in circumstances where explicit knowledge was expected to be used.

In summary, the language background analyses suggested that greater foreign language experience may have been associated with greater acquisition of explicit knowledge. Kemp (2001) and Williams and Lovatt (2005) have reported similar findings. Firstly the rule detectors had reached significantly higher ability levels in their best foreign language than had their non-detector counterparts. Secondly the regression analyses indicated that a rule detector's ability level influenced performance in the untimed test but not in the timed version. Thus, the factor was only relevant for the participants assumed to have explicit knowledge and in the absence of time constraints when that knowledge was predicted to be available. There was no parallel relationship between prior experience and performance in circumstances where performance was assumed to depend on implicit knowledge. Nevertheless, as the experiments did not all produce evidence of implicit knowledge, the non-detector group may not have been homogenous. Instead, it contained participants with accurate implicit knowledge as well as those without. Thus, the seeming irrelevance of language background may have been the result of it not influencing the participants without any relevant implicit knowledge. Nevertheless, the language background analyses highlight a possible additional contrast between performance based on implicit and explicit knowledge, that should be examined more directly in future research. Finally it is reassuring that the development of implicit knowledge, the phenomenon of interest, may not have been affected by the confounding language background variables.

9.3 Reaction Times

Explicit knowledge is assumed to take longer to access than its implicit counterpart (Berry & Dienes, 1993; Bialystok, 1979; Bitan & Karni, 2004). Therefore, the rule detectors who were believed to have explicit knowledge may have taken longer to respond during the grammaticality judgement tests than the non-detectors if the latter had no explicit knowledge rather than inaccurate explicit knowledge. Such a difference is more likely to be detectable in the untimed test when a greater range of response times was possible than in the timed test. Reaction times were calculated separately per group but collapsed across experiments to increase power. (For the purpose of this analysis the rule detectors were combined with the familiarity detectors, as the only relevant feature was whether the participants had developed any explicit knowledge and not which regularity they had actually acquired.)

Figure 9.2 shows that reaction times to correct items in the untimed test were substantially higher than in the timed test. (For the purposes of this analysis, a correct answer was presumed to be a response of *correct* to a grammatical item or of *incorrect* to an ungrammatical item, regardless of whether or not it was also familiar.) An ANOVA with test (timed versus untimed) and group (non-detector versus detector) as independent variables, and reaction time as the dependent variable produced a significant effect of test, $F(1, 122) = 28.675$, $p < 0.001$, $\eta_p^2 = 0.190$, confirming the observation that removing the time constraints did have the desired effect of increasing reaction times. However contrary to the prediction, there was neither a significant main effect of group, $F < 1$, nor an interaction, $F < 1$. Therefore there was no evidence that the non-detectors were responding more quickly in either test than were the rule detectors.

The non-detectors did not respond more quickly than the rule detectors in either the timed or the untimed grammaticality judgement test. Nevertheless, this finding does not contradict the assumption that explicit knowledge takes longer to access. The non-detectors' success in the timed test in Experiments 3 and 4 as compared to the rule detectors demonstrated that the former could respond *accurately* more quickly, and therefore presumably that they could access their knowledge more rapidly. In addition, it is impossible to determine whether the non-detectors were actually using inaccurate explicit knowledge in the untimed tests. If this were the case, there would be no reason to believe that the non-detectors should have responded more quickly.

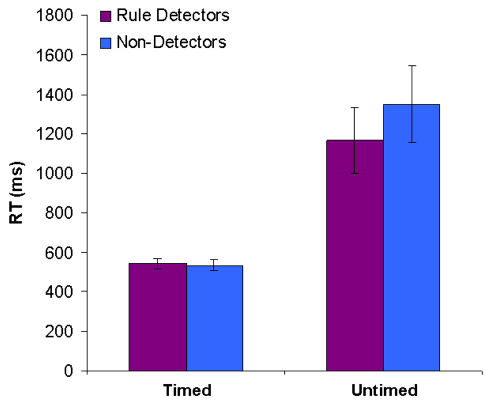


Figure 9.2: Rule detectors' and non-detectors' mean reaction times to the (correctly–answered) judgement test stimuli collapsed across experiments. The error bars show the standard error.

9.4 Chapter Summary

This chapter reported additional analyses of the experimental results, which were predicted to reveal further differences between the acquisition of implicit and explicit knowledge. As expected, higher ability levels in the languages acquired prior to the experiment were found to be associated with the development of explicit knowledge, but not with implicit knowledge. However the reaction time analyses did not produce the predicted outcome. There were no significant differences between the judgement test reaction times of the non-detectors and the rule detectors, who were assumed to be relying mainly on implicit and explicit knowledge respectively.

CHAPTER 10

Simulations

The series of experiments reported in Chapters 4 to 8 suggested that implicit knowledge of one– and two–step dependencies can be acquired (as in Experiments 3 and 4), but that longer ones are not learnt (as in Experiments 1 and 5). A limit is consistent with results in other experimental paradigms (Cleeremans & McClelland, 1991; Johnstone & Shanks, 2001; Newport & Aslin, 2004; Peña et al., 2002), and therefore also with claims that similar processes may be employed in each case. Nevertheless, questions remain that could not be addressed experimentally within the time constraints of the current project.

A series of connectionist simulations was employed to support and extend the experimental findings. They are an effective way of determining what can be learnt from the input statistics, given a certain set of parameters. Therefore they provide hypotheses as to what humans may achieve if they only learnt in the same way. As such, discrepancies between human and network performance indicate circumstances in which additional factors may influence the acquisition of implicit knowledge. On the other hand, similarities between human and network performance are consistent with the interpretation that the acquisition of implicit knowledge is largely driven by the input statistics.

There are many advantages to using simulations in conjunction with experiments when investigating the development of implicit knowledge. First, unlike human participants, a connectionist network is guaranteed to have no previous knowledge of any other language. If it successfully acquires a structure, it does so without the aid of transfer. Second, as a network only simulates implicit knowledge this cannot be confused with automatised explicit knowledge (N. C. Ellis, 2003). Longer exposure periods are therefore possible. Third, as a network does not utilise language–specific learning modes, if it is able to develop knowledge of the linguistic rules it must do so by means of domain–general processes. Fourth, given these constraints, a simulation confirms exactly what can be learnt from the statistics of the input. Finally, the input coding can be manipulated to alter the degree of similarity between different items. This chapter reports simulations that address all of these points.

There has been a successful history of using connectionist networks to simulate experimental findings of implicit learning (see N. C. Ellis, 2003, for a review). As previously described, Cleeremans and McClelland (1991) trained participants in the serial response task. When they modelled the results with a Simple Recurrent Network (SRN) the human data pattern was successfully reproduced. In Dienes, Altmann, Gao, and Goode (1995) participants were exposed to the output of a finite state grammar and were then tested either in the same or a different (transfer) modality. The transfer participants performed better than untrained controls, but they were at a disadvantage compared to their same modality counterparts. An SRN was shown to reproduce both sets of human data. Dienes, Altmann, and Gao (1999) used SRNs to simulate a variety of previous experimental studies into artificial grammar learning. They replicated the results of Brooks and Vokey (1991) using a network that became sensitive both to the grammaticality of test strings and to their similarity to the training items. Their model reproduced the findings of successful transfer performance when letters were repeated in the strings as originally reported by Whittlesea and Dorken (1993) and similar findings of transfer in strings without repetition reported by Altmann et al. (1995). They also confirmed that a network, like the human participants tested by Gómez and Schvaneveldt (1994), learnt more from exposure to complete strings than to individual bigrams. Finally, they were also able to extend the last set of findings by demonstrating that when the number of bigrams in the input of the two exposure conditions was equalised, the contrast disappeared. Thus, connectionist simulations have not only been used to replicate numerous experimental findings in the implicit learning literature, but they can also extend the results. Nevertheless, rather than firm conclusions about human performance, the outcomes should be treated as hypotheses for future experiments.

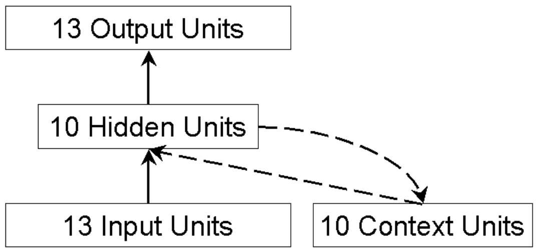


Figure 10.1: Simple recurrent network used in the simulations.

10.1 Method

10.1.1 Architecture

A standard SRN as in Figure 10.1 was employed to model and extend the experimental results from Chapters 4 to 8 using the Light, Efficient Network Simulator (LENS) software package[1] (Rohde & Plaut, 1999). These networks contain an input, a hidden, a context and an output layer. Activation proceeds from the nodes in the input to the hidden to the output layer. At each time–step, the values of the units in the hidden layer are also copied to the context layer. They are then fed back into the hidden layer at the same time as the next set of input values. This allows the network to take context into account, and makes SRNs ideally suited for the acquisition of sequences (Elman, 1993). Each node in the hidden layer was connected to every node in both the input and output layers. The initial weights on those connections were set randomly within the default range assigned by the program.

There were thirteen units in both the input and output layers, and ten units in the hidden and context layers. The input was encoded using a localised representation. One input and one output unit were used as a boundary marker between strings while the remaining twelve represented the target domain. Table 10.1 shows the specific encoding used in each simulation. Simulation 1 modelled the biconditional grammar experiment using two nodes for each of the six letters involved. In the majority of the linguistic simulations, one unit corresponded to each morpheme. Two of the units were unused in Simulations 2 and 3a. However in the word–based encoding used in Simulation 3b (see Section 10.2.3 below), a single node represented each fully inflected word form. Case was not encoded in the Basque stimuli.

[1]Parameters not specified here received the default values that were assigned by the program

	Biconditional	Persian Morph.	Persian Word	Basque
Unit 1	D	Mehmun	Mehmun	Motilye\a
Unit 2	D	Shohar	Mehmunan	Gisone\a
Unit 3	F	Yateem	Shohar	Lahune\a
Unit 4	F	Geer	Shoharan	Ekarri
Unit 5	G	Kesh	Yateem	Ahurtu
Unit 6	G	Tars	Yateeman	Ikusi
Unit 7	K	-Ø (sing. noun)	Geere	-Ø (sing. noun)
Unit 8	K	-an (plural noun)	Geerad	-k (plural noun)
Unit 9	L	-e (sing. verb)	Keshe	Du-
Unit 10	L	-ad (plural verb)	Keshad	Deetu-
Unit 11	X		Tarse	-Ø
Unit 12	X		Tarsad	-te
Unit 13	Boundary	Boundary	Boundary	Boundary

Table 10.1: Encoding used in the connectionist simulations. Experiment 3 was simulated in two ways: with one unit encoding each Persian morpheme and with one unit encoding each inflected Persian word.

10.1.2 *Procedure*

The network was trained on a next–item prediction task using back–propagation of error. Strings or sentences were input into the network one element at a time with the relevant input nodes given an activation level of one, and the remainder set at zero. The model's task was to predict the next letter or word in the string or sentence, by producing the same activation pattern in the output units that would be introduced to the input layer in the following time step. The error in this prediction was used to adjust the weights using the *steepest descent* option in LENS. The learning rate was 0.05 throughout. As was the case with the other parameters, this was set by piloting Simulation 3a (that used the Persian materials with the source localisation task) to maximise similarity to the human data. This simulation was selected for piloting because the corresponding experiment produced evidence of significant implicit knowledge. Importantly, the parameters were then used unchanged for the remainder of the simulations.

One single network trial was used to model each experimental trial both in the exposure phase and in the test, as recommended by Dienes et al. (1999). Two lengths of simulations were employed. A standard exposure phase was the equivalent of the experimental one, containing the same number of trials (that is, eight repetitions of each learning string). An extended version employed sixty thousand learning trials, the same number used by Cleeremans and

McClelland (1991) to investigate the maximum learning outcomes for long–distance dependencies. In this version, there were two thousand five hundred repetitions of each learning string (three hundred and twelve and a half times the number in the standard version). To keep the proportion of tokens of each type the same as in the corresponding experiment, the number of exposures to each reproduction/source localisation task question item was scaled similarly. Thus, each one was repeated either three hundred and twelve or three hundred and thirteen times.

The network's task was always next–word prediction, regardless of whether a simulation was modelling the reproduction or the source localisation task. For the model the only difference between the learning tasks was the input. There were additional repetitions of individual learning items in the source localisation task, modelling the input provided by the source–localisation questions themselves. For the reproduction task the human learners' attempts to type the stimuli were interspersed with the learning items. In the experiments there was only one source localisation question for every four exposure trials, while every single trial in the reproduction task was followed by a question. Thus the network trained on the reproduction task received more trials, but many of them involved incorrect input.

The participants' reproduction task responses in Experiment 2 were converted into input vectors for the simulation as follows. Spelling errors were ignored, as long as the intended target could still be identified. For example, *yateemen* was encoded as if the target *yateeman* had been produced. However morphemes that could not be uniquely identified were omitted. Therefore a response such as *kesha* that was midway between the singular form *keshe* and the plural form *keshad* was encoded simply as *kesh*, with neither the singular nor the plural unit activated. The resulting input vectors included the participants' errors in morpheme selection and order, but not in orthography.

One run of the network simulated each non-detector or control participant. The rule and familiarity detectors were not included as the aim was to simulate the acquisition of implicit knowledge. The items were presented to the network in the same order as they had been to the corresponding experimental participant. Thus, the order of presentation was different for each run.

The simulation only modelled the learning phase and the timed test. Neither the untimed nor the correction tests were included, as both were assumed to measure performance based largely on explicit knowledge. The weights were not frozen during the test, following from the assumption that implicit knowledge develops automatically and continuously. As in the learning phase, the network completed the next–word prediction task during the test and it was trained according to back–propagation of error.

10.1.3 Analysis

To make the simulation results comparable with the human data, the network's output was converted into grammaticality judgements and analysed using signal detection theory. There is more than one way to achieve this. Marcus (1999b) recommended calculating the number of morphemes that were predicted correctly. However when this was attempted with the current results the output was identical for many items. As a result it was impossible to select a threshold whereby a sufficient balance between the number of classifications of *correct* and *incorrect* allowed accurate d' values to be obtained. An alternative calculation was employed by Dienes (1992) and Dienes et al. (1999), based on the cosine of the angle between the actual and predicted activation vectors. However this is calculated according to the formula in (22*a*), and involves multiplying the two vectors together to find the dot product as in (22*b*), where *a* is the target activation vector and *b* the actual output vector (James et al., 1996). As the inactive nodes in the target vectors were set to zero, such a calculation would remove any influence of error in which output nodes were active that should not have been. Instead, it would depend solely on error in which activation levels were lower than the target. Instead of these calculations therefore, the Euclidean distance between the output and target activation vectors was calculated for each test item following Altmann and Dienes (1999). The value for each stimulus was then compared to the median for the relevant block of the test in that run of the simulation. A Euclidean distance lower than the median was interpreted as a classification of *correct*, while a higher one was equated to an *incorrect* classification. Note that the median was calculated separately for each of the two test blocks, as the response bias analyses in Sections 4.2.4, 6.2.5 and 7.2.4 suggested that the participants relying on implicit knowledge tended to adjust their threshold as the experiment progressed.

(22) a) $cos\theta = \frac{a \cdot b}{|a| \times |b|}$

b) $a \cdot b = a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 + a_5b_5 + a_6b_6 + a_7b_7 + a_8b_8 + a_9b_9 + a_{10}b_{10} + a_{11}b_{11} + a_{12}b_{12} + a_{13}b_{13}$

Once the network's performance was expressed as grammaticality judgements it was possible to calculate each run's sensitivity to both grammaticality and familiarity (d') using SDT. The simulation d' values were compared to untrained control runs of the network in the same way as previously, firstly using ANOVAs with condition (trained versus control network) as a between–subjects factor and block as a within–subjects factor. If the omnibus analyses revealed a significant or marginal effect involving the variable block one–tailed t-tests were then used for planned comparisons between the trained and control networks separately for each block. Alpha was set to $p < 0.05$ for the ANOVAs, and reduced to $p < 0.025$ for the t-tests. Effects significant at the $p < 0.1$ level in the ANOVAs ($p < 0.05$ in the t-tests) were reported as marginal.

The simulation was assumed to be modelling the acquisition of implicit knowledge (N. C. Ellis, 2003). However two of the criteria used to confirm this were not applicable to the simulations. There was no time constraint in the judgement test and the correction test was not included in the procedure. However the remaining criterion was still assessed. Performance reflecting the equivalent of implicit knowledge was assumed to be significant in block one, but not in block two.

10.2 Results

10.2.1 Simulation 1

The participants in Experiment 1 were exposed to the output of a biconditional grammar, where the target structures were four–step dependencies. The non-detectors acquired sensitivity to the frequent bigrams that they demonstrated throughout the test, but they were never sensitive to the grammaticality of the test strings. This experiment was modelled in two ways: firstly with an exposure length equivalent to the experiment, and secondly with an extended period to approximate the maximum learning outcomes. In both cases there were ten trained runs of the network to model the ten non-detectors, and sixteen control runs modelling the sixteen control participants.

Simulation 1a: Standard Learning Phase

Simulation 1a employed a standard learning phase comparable to Experiment 1. (A sample of the input is provided in Appendix G.) A close match with the

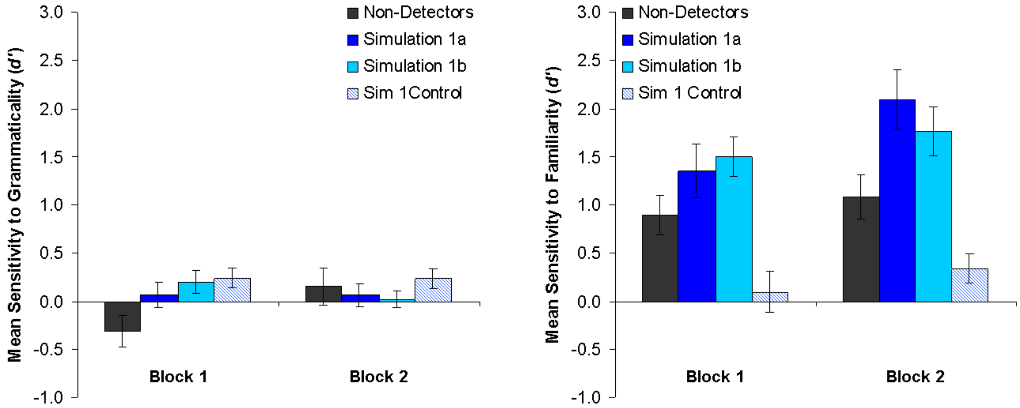


Figure 10.2: Simulation 1 results. The figure shows the mean sensitivity to each of the dependent variables as measured in d' units for the trained and control runs of the network. The non-detectors from Experiment 2 are included for comparison. Error bars show the standard error.

experimental results would therefore support the use of such models. Figure 10.2 shows the experimental and simulation results for comparison. The trained runs were less sensitive to grammaticality in both blocks than were the control networks. As the trends were in the opposite direction to that predicted by the one–tailed hypothesis, no statistical tests were carried out.

Figure 10.2 clearly suggests that the network became sensitive to the frequent bigrams and that it remained so throughout the test. An ANOVA with block as a within–subjects factor, condition (trained versus control) as a between–subjects factor and sensitivity to familiarity as the dependent variable revealed a main effect of condition, $F(1, 24) = 50.062$, $p < 0.001$, $\eta_p^2 = 0.676$. There was neither a main effect of block, $F(1, 24) = 2.140$, ns, $\eta_p^2 = 0.082$, nor a block x condition interaction, $F < 1$. Therefore the blocks were not analysed separately.

In summary, the network did not learn the target four–step dependencies, but it was sensitive to familiarity throughout the test. This result mirrored Experiment 1, where contrary to the predictions for implicit knowledge, the non-detectors remained sensitive to familiarity throughout the judgement tests. The simulation confirmed that familiar and unfamiliar strings could be distinguished purely from the statistics of the input, and that this contrast remained even after novel bigrams had been introduced in the timed test. Thus, as the non-detectors' sensitivity to familiarity in Experiment 1 followed the input statistics, it may have been the result of implicit knowledge. Nevertheless firm conclusions could not be drawn.

Simulation 1b: Extended Learning Phase

Neither the non-detectors nor the network learnt the four–step dependencies during the standard exposure period. A second simulation using sixty thousand learning trials therefore tested whether the structure could be learnt with more exposure. According to Figure 10.2 however, the network's sensitivity to grammaticality was still not significantly above chance, with the trend in the opposite direction to that predicted throughout.

As in the shorter version, an ANOVA with sensitivity to familiarity as the dependent variable resulted in a significant effect of condition, $F(1, 24) = 117.841$, $p < 0.001$, $\eta_p^2 = 0.831$, but neither a significant main effect of block, nor a block x condition interaction, all Fs < 1. Thus, extending the learning phase did not alter the findings. The network did not learn the target structure, but it remained sensitive to the frequent bigrams even after exposure to novel versions. Overall therefore the results are consistent with the interpretation that implicit knowledge of four–step dependencies cannot be acquired from the statistics of the input. Nevertheless, this prediction should be confirmed experimentally.

10.2.2 Simulation 2

In Experiment 2, the participants were exposed to Persian input with subject–verb agreement as the target structure, instantiated as a two–step dependency. They were trained using the reproduction task. Under these circumstances the non-detectors did not acquire implicit knowledge either of the target two–step structure or of the frequent bigrams. A corresponding simulation therefore tested whether the target structure could be acquired from the input statistics, or whether the difficulty was caused by the type of processing required in the learning task. There were twenty–four trained runs modelling the twenty–four non-detectors, and thirty–six control runs modelling the same number of control participants.

Figure 10.3 shows the trained and control runs' sensitivity to both grammaticality and familiarity, together with that of the non-detectors for comparison. An ANOVA with condition (trained versus control) and block (one versus two) as independent variables and sensitivity to grammaticality as the dependent variable indicated that there were no significant main effects of condition or block, nor an interaction between the two factors, all Fs < 1. A corresponding ANOVA with sensitivity to familiarity as the dependent variable once again did not produce a main effect of condition, $F(1, 58) = 1.703$, ns, $\eta_p^2 = 0.029$, or of block, $F < 1$,

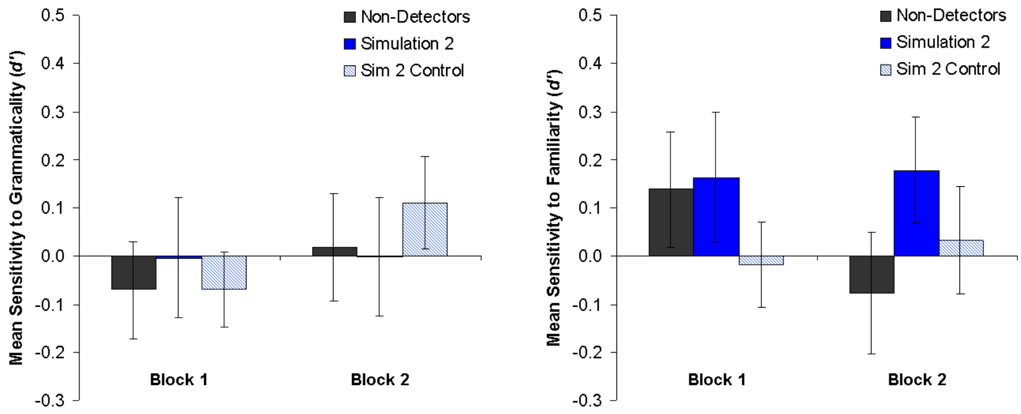


Figure 10.3: Simulation 2 results. The figure shows the mean sensitivity to each of the dependent variables as measured in d' units for the trained and control runs of the network. The non-detectors from Experiment 2 are included for comparison. Error bars show the standard error.

or an interaction, $F < 1$. In summary, Simulation 2 was consistent with the experimental results, indicating that neither the target structure nor the familiarity manipulation was learnable. Thus, whatever aspect of the reproduction task prevented the development of implicit knowledge in Experiment 2 may have been shared by the simulation.

10.2.3 Simulation 3

Simulation 3a: Morphemic Encoding

When the experimental learning task was changed from reproduction to source localisation for Experiment 3, the new group of twenty–three non-detectors acquired the target two–step dependency but still not the frequent bigrams. Their performance with the former met the three criteria to be classified as the result of implicit knowledge. Their performance was modelled using twenty–three runs of the network and contrasted with the control runs from Simulation 2.

Figure 10.4 shows the output of the corresponding simulation. An ANOVA with condition as a between–subjects factor and block as a within–subjects factor was conducted on the network's sensitivity to grammaticality. It did not reveal significant main effects of condition, $F(1,57) = 1.821$, ns, $\eta_p^2 = 0.031$, or of block, $F < 1$. However there was a marginal interaction between the two factors, $F(1,57) = 3.167$, $p = 0.08$, $\eta_p^2 = 0.053$. Subsequent planned comparisons revealed a significant effect of condition in block one, $t(57) = 2.186$, $p = 0.0165$, $\eta_p^2 = 0.077$, while Figure 10.4 shows that the trend was in the opposite direction to that predicted in block two. A parallel ANOVA with sensitivity to familiarity as the dependent variable demonstrated that there were no significant main effects of condition, $F(1, 57) = 1.297$, ns, $\eta_p^2 = 0.013$, or of block, $F(1, 57) = 1.114$, ns, $\eta_p^2 = 0.019$, and nor was there a condition x block interaction, $F < 1$.

Like the non-detectors, the network acquired sensitivity to the target structure and its performance became less accurate following exposure to the incorrect judgement test stimuli. It is not surprising that the network mirrored the human results however, because the parameters were originally set by piloting Simulation 3a to maximise similarity with the human data. The contrast between the results of Simulations 2 and 3a suggests that the incorrect input provided by the reproduction task was sufficient to prevent implicit learning of a target structure that could otherwise be acquired. Nevertheless it cannot prove that the incorrect input rather than another aspect of the reproduction task caused the difficulties for the human participants in Experiment 2.

As in Experiment 3, it was also necessary to carry out an attraction error analysis to assess whether the network had acquired sensitivity to the target two–step dependency between the number of the first noun (the possessee) and that of the verb, or whether it had acquired the confounding one–step dependency between the number of the second noun (the possessor) and that of the verb. This possibility was the result of an imbalance in the stimuli, explained above in Section 6.2.6. When the number value of the two nouns was the same, in the match items, the network was expected to be sensitive to grammaticality. However when one of the nouns was singular and the other plural in the mismatch items, a network that had acquired the two–step target structure would be expected to demonstrate sensitivity to grammaticality, whereas one that had learnt the one–step confounding dependency would be *less* sensitive to grammaticality than were the control networks.

The results of the attraction error analysis can be seen in Figure 10.5. Separate ANOVAs were conducted for the number–match and number–mismatch items, with condition and block as independent variables and sensitivity to grammaticality as the dependent variable. ANOVAs producing significant or marginal effects of block or interactions involving the variable were followed up with separate two–tailed t-tests for each block, with alpha reduced to $p < 0.025$.

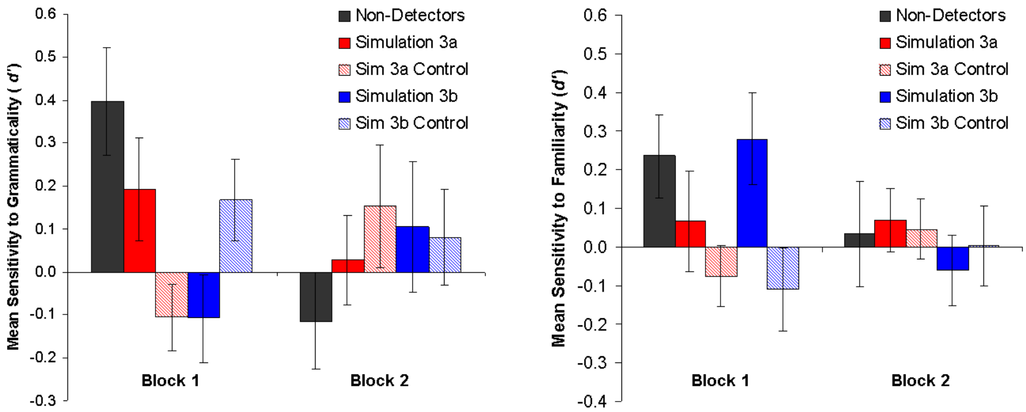


Figure 10.4: Simulation 3 results. The figure shows the mean sensitivity to each of the dependent variables as measured in d' units for the trained and control runs of the network. The non-detectors from Experiment 2 are included for comparison. Error bars show the standard error.

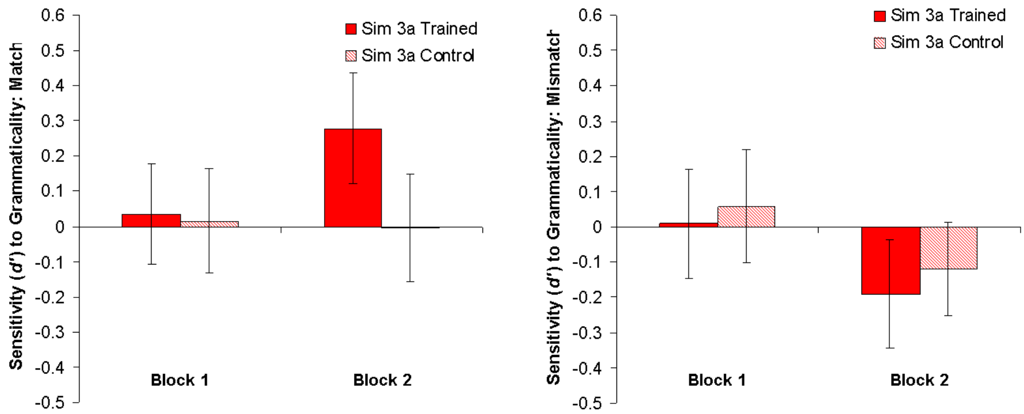


Figure 10.5: Mean sensitivity to grammaticality in Simulation 3 in the number–match and number–mismatch items, as measured in d' units. The error bars show the standard error.

The ANOVA for the mismatch items did not reveal a main effect of block, $F(1, 57) = 1.612$, ns, $\eta_p^2 = 0.027$, or one of condition, $F < 1$, and nor was there a significant interaction, $F < 1$. The same was true of the ANOVA for the match items, all Fs < 1.

As the trained networks were not less sensitive to grammaticality in the mismatch items than were the control networks, the attraction error analysis did not provide any evidence that the networks had acquired the confounding one–step dependency between the number value of the second noun and that of the verb. Rather, the effect demonstrated in the main analyses is likely to have been the result of knowledge of the target two–step dependency.

Simulation 3b: Word Encoding

Neither the network in Simulations 2 or 3a nor the participants in Experiments 2 or 3 acquired the frequent bigrams in the Persian materials implicitly. Three explanations for the finding have been suggested. First, the ACS difference between the familiar and unfamiliar items was smaller than in the other sets of materials. The difference between the Persian and biconditional materials was confirmed statistically in Section 5.3, while that between the Persian and the Basque sets was not tested because the variance was zero in the latter case. Second, the majority of the unfamiliar items were identified as such by the fully inflected words involved. Thus, *yateem tarsad* could appear in the learning items, while *yateem tarse* would still be considered a novel bigram in the test phase. If the participants did not sufficiently distinguish between these similar forms, they could not have demonstrated sensitivity to familiarity regardless of the size of the ACS difference. Third, the imbalance in the numbers of familiar and unfamiliar items may have made knowledge difficult to detect. (A fourth option, that language learners are biased against acquiring regularities based on individual lexical items rather than on word–classes was ruled out by Experiment 4.)

Simulation 3b addressed the second option. In the previous simulation each morpheme was encoded separately, so that words that shared morphemes also shared active input nodes. Thus, the divide between inflected and lexical familiarity was maintained. In contrast in Simulation 3b each inflected word was encoded as a separate input unit. Following this change, words containing the same lexical root but different affixes were no more similar to each other than were words built from different roots. This removed the inflected familiarity option, so that all test items were either familiar or lexically unfamiliar. If implicit knowledge of lexical but not inflected familiarity can be acquired from this set of materials, then its effects should emerge in the current simulation. As the change in encoding affected the test and not just the learning phase, new control runs of the simulation were also needed.

Following the change to lexical familiarity, an ANOVA with condition (trained versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable did not reveal any significant effects, block $F < 1$, condition, $F(1, 57) = 1.131$, ns, $\eta_p^2 = 0.019$, condition x block interaction, $F(1, 57) = 1.624$, ns, $\eta_p^2 = 0.028$. Thus, the network did not acquire the target structure. A corresponding ANOVA with sensitivity to familiarity as the dependent variable did not uncover a significant effect of condition, $F(1,57) = 2.080$, ns, $\eta_p^2 = 0.035$, nor one of block, $F(1,57) = 1.062$, ns, $\eta_p^2 = 0.018$. However there was a significant condition x block interaction, $F(1,57) = 4.246$, $p = 0.044$, $\eta_p^2 = 0.069$. Planned comparisons produced a significant effect of condition in block one, $t(57) = 2.357$, $p < 0.011$, $\eta_p^2 = 0.089$ with the trained networks showing greater sensitivity to familiarity than the controls. The effect was not significant in block two, $t < 1$, thereby fitting the criterion for classification as the result of implicit knowledge.

Unlike in Simulation 3a, the network did not successfully model the non-detectors' sensitivity to grammaticality, and unlike both the previous simulation and Experiment 3 it did learn the frequent bigrams. The simulation results are consistent with the interpretation that the non-detectors' inability to acquire the frequent bigrams in the Persian materials was caused by the use of inflected familiarity. This conclusion runs counter to the results of the type–of–familiarity analysis in Section 6.2.4, which found that the non-detectors were sensitive neither to inflected nor to lexical familiarity individually. Either a lack of power may have prevented the participants' sensitivity to lexical familiarity from reaching significance, or the human learners and the network may differ in this regard. Interestingly however, when the network did acquire the frequent bigrams in Simulation 3b it was no longer able to learn the target two–step dependency within the short exposure period. This suggestion, that the acquisition of longer structures is prevented by the acquisition of shorter types, is relevant to the interpretation of Experiment 4 in which the non-detectors became sensitive to familiarity but surprisingly not to the target two–step dependency.

10.2.4 Simulation 4

Simulation 4a: Standard Learning Phase

The non-detectors in Experiment 4 acquired implicit knowledge of the frequent bigrams, performing above chance in block one of the timed test but not in block two. However, in contrast to those exposed to subject–verb agreement in Experiment 3, the group did not learn the target two–step dependency representing object–verb agreement. As a result, we proposed that the learners may have acquired the abstract categories in order to be sensitive to this difference, and non–local dependencies may have required transfer from a previous language.

On the other hand the network learnt the two–step dependency in Simulation 3 successfully without using transfer. If a network can acquire long–distance dependencies without transfer but human learners rely on it, then the current simulation of Experiment 4 should reveal sensitivity to the target structure. However if a different aspect of the input in Experiment 4 (such as the presence of a learnable contrast between familiar and novel bigrams) prevented successful learning, then the same factor may impact on the current network's performance. There were seven trained runs of the network to model the seven non-detectors, and eighteen control runs to model the control participants.

The results for Simulation 4a can be seen in Figure 10.6. An ANOVA with condition (trained versus control) and block (one versus two) as the independent variables and sensitivity to grammaticality as the dependent variable suggested that there were no main effects of condition, $F < 1$, or block, $F(1, 23) = 1.049$, ns, $\eta_p^2 = 0.044$, nor an interaction, $F < 1$. Thus, there was no evidence that the network acquired the target structure. A parallel ANOVA based on sensitivity to familiarity produced a significant effect of condition, indicating that there was a difference between the trained and control runs of the network, $F(1, 23) = 8.433$, $p = 0.008$, $\eta_p^2 = 0.268$. There was also a significant main effect of block, $F(1, 23) = 5.902$, $p = 0.023$, $\eta_p^2 = 0.204$ and a condition x block interaction, $F(1, 23) = 14.187$, $p = 0.001$, $\eta_p^2 = 0.381$. Planned comparisons confirmed that, as expected from implicit knowledge, the effect was significant in block one, $t(23) = 4.521$, $p < 0.001$, $\eta_p^2 = 0.471$, but not in block two, $t < 1$. Again, the simulation provided a close fit to the human data and it fitted the criterion of implicit knowledge.

The simulations replicated the human performance both in Experiment 3, where implicit knowledge of two–step subject–verb agreement was acquired, and in Experiment 4 where an equally long object–verb agreement structure was not learnt. Previously an explanation based on transfer was posited for the difference between the two linguistic structures in the human data. However the network followed the same pattern, without any mechanism for transfer or any previous linguistic knowledge, suggesting that the results could be explained without claiming that the participants were sensitive to the abstract linguistic structures. The consequences of this finding are considered in more detail in the following section.

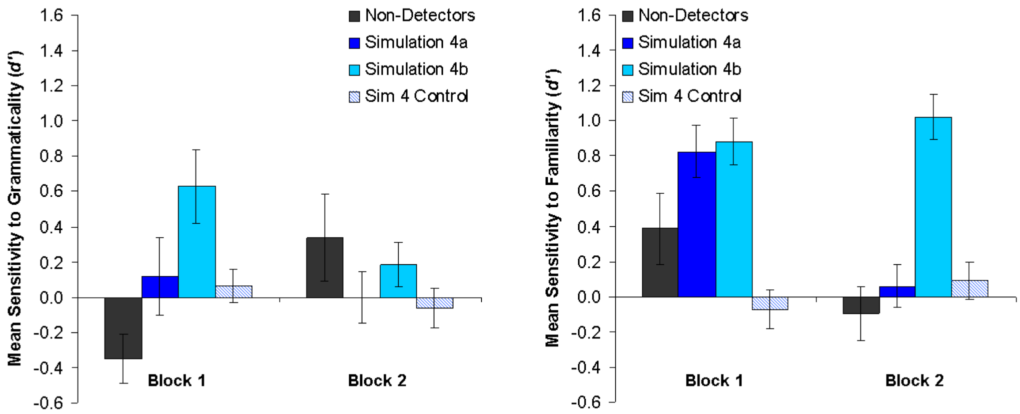


Figure 10.6: Simulation 4 results. The figure shows the mean sensitivity to each of the dependent variables as measured in d' units for the trained and control runs of the network. The non-detectors from Experiment 2 are included for comparison. Error bars show the standard error.

Simulation 4b: Extended Learning Phase

The participants in Experiment 4 and the network in Simulations 3b and 4a acquired the frequent bigrams but not the longer target structures. The opposite was the case in Experiment 3 and Simulation 3a. Implicit knowledge of both structures has not yet been found in any one experiment or simulation in this thesis. Thus, learning the frequent bigrams may have prevented or delayed the acquisition of longer patterns so that it was not possible to achieve both within the confines of the procedure. Therefore it is possible that the two–step dependency in Experiment 4 may have been acquired given a longer learning phase. Simulation 4b tested this prediction, extending the exposure period to sixty thousand trials.

The network's sensitivity to each independent variable can be seen in Figure 10.6. An ANOVA with sensitivity to grammaticality as the dependent variable indicated that there was a significant effect of condition, $F(1, 23) = 7.779$, $p = 0.010$, $\eta_p^2 = 0.253$, but not one of block, $F(1, 23) = 1.903$, ns, $\eta_p^2 = 0.076$, and no interaction, $F < 1$. The same was true when sensitivity to familiarity was the dependent variable, condition $F(1, 23) = 12.854$, $p = 0.002$, $\eta_p^2 = 0.359$, all other Fs < 1.

Once the learning phase had been extended, the network acquired both the target two–step dependency and the frequent bigrams, and it remained sensitive to both throughout the procedure. The simulation suggests that, if the human participants learnt via a statistical tallying process, they may have done the same in an extended experiment. This hypothesis should be tested in future experiments, although different ways of distinguishing implicit and explicit knowledge may be required.

10.2.5 Simulation 5

Simulation 5a: Standard Learning Phase

The participants in Experiment 5 did not develop implicit knowledge of either the target three–step dependency or the frequent bigrams. The former was not surprising as there is little evidence from other sources that implicit knowledge of three–step dependencies is acquired. However the bigrams result ran counter to the predictions. The ACS contrast was numerically larger than it was in the materials for Experiments 1 and 4, where sensitivity to the pattern did develop. (The significance of these differences was not tested statistically because the variance within each item type in Experiment 5 was zero.) Therefore a simulation was used to confirm whether there was indeed a sufficient difference between the familiar and unfamiliar items for learning to occur based on the input statistics. There were eight trained runs to model the non-detectors and eighteen untrained control runs to model the control participants.

The results are shown in Figure 10.7. An ANOVA with condition (trained versus control) and block (one versus two) as independent variables and sensitivity to grammaticality as the dependent variable revealed that there were no significant main effects of condition or block, all Fs < 1, but that there was a marginal interaction between the two, $F(1, 24) = 3.838$, $p = 0.062$, $\eta_p^2 = 0.138$. In block one the trend was in the opposite direction to that predicted, while a planned comparison demonstrated that the effect of condition was not significant in block two $t(24) = 1.525$, ns, $\eta_p^2 = 0.088$. Overall there is no evidence that the network acquired the target three–step dependency, consistent with the non-detectors in Experiment 5.

A parallel ANOVA on sensitivity to familiarity did not indicate a significant effect of condition, $F(1, 24) = 1.344$, ns, $\eta_p^2 = 0.053$, but there was a marginal effect of block, $F(1, 24) = 3.117$, $p = 0.090$, $\eta_p^2 = 0.115$, and a significant interaction between the two variables, $F(1, 24) = 8.081$, $p = 0.009$, $\eta_p^2 = 0.252$. Planned comparisons revealed that the trained networks were significantly more sensitive to the variable in block one than were the untrained controls, $t(24) = 2.325$, $p = 0.0145$, $\eta_p^2 = 0.184$. However the effect had disappeared by block two, where the trend was not in the predicted direction. Unlike the human participants,

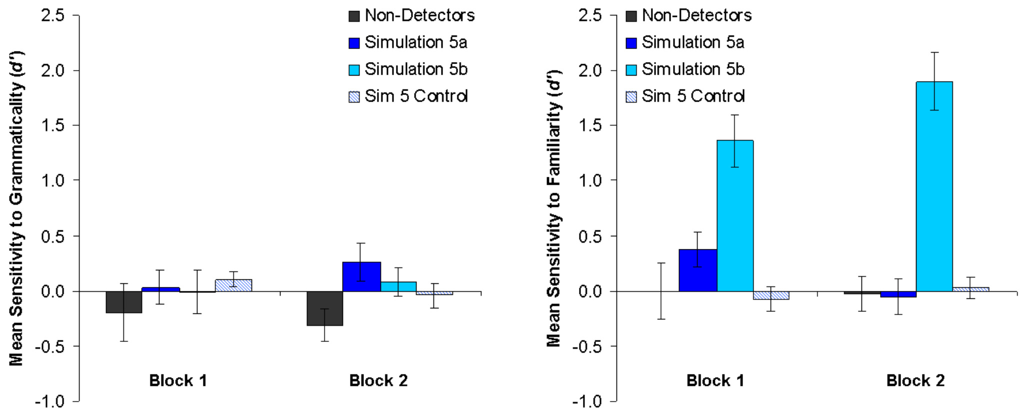


Figure 10.7: Simulation 5 results. The figure shows the mean sensitivity to each of the dependent variables as measured in d' units for the trained and control runs of the network. The non-detectors from Experiment 2 are included for comparison. Error bars show the standard error.

the network had acquired the frequent bigrams and its performance fitted the criterion for implicit knowledge. Thus the simulation confirmed that the structure could be acquired from the statistics of the input. An alternative explanation is therefore required for the participants' performance.

Simulation 5b: Extended Learning Phase

The two–step dependency in Simulation 4 was not acquired in a standard learning phase equating to that used in the experiments. However it was once the learning phase had been extended. The same may apply for the current target structure. Therefore a similarly lengthened version of Simulation 5 was used to assess whether implicit knowledge of three–step dependencies could be acquired with sixty thousand exposure trials.

An ANOVA with sensitivity to grammaticality as the dependent variable did not reveal a main effect of condition, $F < 1$, one of block, $F < 1$, or an interaction, $F(1, 24) = 1.105$, ns, $\eta_p^2 = 0.044$. Even following the extended learning phase therefore, there was no evidence that the network had acquired the target three–step dependency. Thus, if human learners rely solely on the statistics in the same way as the network they would be unlikely to acquire this feature. Supporting experimental evidence from human performance is required in order to examine this possibility.

A parallel ANOVA demonstrated that there was a significant effect of condition on sensitivity to familiarity, $F(1, 24) = 107.007$, $p < 0.001$, $\eta_p^2 = 0.817$, and a marginal effect of block, $F(1, 24) = 4.075$, $p = 0.055$, $\eta_p^2 = 0.145$, but not a significant interaction between them, $F(1, 24) = 1.885$, ns, $\eta_p^2 = 0.073$. Planned comparisons showed that the effect of condition was significant both in block one, $t(24) = 6.310$, $p < 0.001$, $\eta_p^2 = 0.624$, and in block two $t(24) = 8.459$, $p < 0.001$, $\eta_p^2 = 0.749$. Following the extension to the learning phase, the trained networks were sensitive to the frequent bigrams and they remained so throughout. This is unsurprising because the ratio of familiar to novel bigrams was larger than in the shorter version, and therefore the exposure to the latter in the test phase was unlikely to have been sufficient to alter the network's representation of the target domain substantially.

10.3 Discussion

In many cases there was a close match between the experimental data and the simulations with an equivalent exposure period. The participants and the networks trained on the biconditional grammar were both sensitive to the frequent bigrams and only to the bigrams throughout the testing period (which corresponded to the timed test in the experiments). Those exposed to the version of the Persian stimuli that utilised inflected familiarity and trained in the source localisation task only learnt the target rule, and their knowledge was only measurable in block one of the test. Both the network in Simulation 4a and the participants in Experiment 4 were sensitive only to the frequent bigrams, and again only in the first block of the test. Specifically, the simulations replicated the length limit on the acquisition of implicit knowledge suggested by the experiments. Given sufficient exposure time, the network became sensitive to both of the two–step dependencies but never to the single three–step contingency or to the multiple four–step versions. These parallel results suggest that the particular network was a good model of the participants' performance over the course of the current experiments, and therefore endorse its use to derive predictions of human behaviour in different situations.

What a specific connectionist network learns is entirely dependent on the statistical distribution of the input and its architecture. The close correspondence between the experimental and simulation data therefore suggests that statistical learning is likely to contribute to the acquisition of implicit knowledge in human learners. This conclusion is also supported by the ACS values. When the non-detectors became sensitive to the frequent bigrams in Experiment 4, the effect was only measurable in block one of the timed test. However the equivalent in Experiment

1 was maintained throughout the judgement tests. Both effects could be predicted by the changes in ACS values. The familiar and unfamiliar items in Experiment 1 could still be distinguished in this way at the end of the timed test, but the same was not true for Experiment 4 (see Table 4.1 in Section 4.1.2 above). Nevertheless although such learning may occur in adult SLA, second language acquisition in natural environments is unlikely to be restricted to statistical tallying.

Neither the network in Simulation 3a nor the participants in Experiment 3 distinguished the familiar and unfamiliar strings in the Persian data. However, once the inflected familiarity manipulation had been replaced by lexical familiarity in Simulation 3b, the network became sensitive to the pattern. In contrast to the type–of–familiarity analysis in Section 6.2.4 therefore, the simulation suggested that frequent and novel bigrams could only be distinguished when they include different lexical roots and not only different inflections. As the type–of–familiarity analysis was based on so few items, the simulation results may be of greater importance.

The network became sensitive to the target subject–verb agreement in Simulation 3a when the source localisation task was modelled, but not when the reproduction task was modelled in Simulation 2. The same was true of the participants. From the network's point of view the only difference between the two versions was in the input. Networks trained on the reproduction task received the human participants' (often incorrect) attempts to retype the stimuli in addition to the designated learning items. Thus, the incorrect input was sufficient to cause the learners' problems. However it may not have been the sole origin of the participants' difficulties. Their reproductions of the learning phase stimuli were not displayed on the computer monitor but were probably represented phonologically in short–term memory, thereby functioning as self–generated input. Nevertheless, a different aspect of the reproduction task, such as the requirement for detailed rather than holistic processing, may have contributed to the experimental findings. Any assessment of which aspects of the task were responsible remains speculative and a detailed discussion of the point is outside the realm of this thesis.

The results of Experiment 5 and Simulation 5a differed in regard to familiarity. Surprisingly, the human participants were not sensitive to the frequent bigrams despite the strength of the ACS contrast. However the network was, even after the same amount of exposure. While the results of Simulation 5b suggests that performance would improve following a longer learning phase, it is important to consider why the participants and the network differed following the standard amount of exposure. Section 8.3 offered a tentative explanation of the learners' performance based on attention. This would be consistent with the network's performance because attention was not encoded or manipulated in the simulation. Nevertheless (an)other aspect(s) of the materials or procedure may have contributed to the findings.

The success of Simulation 4b with the target two–step object–verb agreement suggested that the participants may have done the same if they had been given more exposure. The network learnt the structure without transferring knowledge from a previous language, thereby demonstrating that the statistics of the input were sufficient for learning to take place. Thus, the result suggested that the contrast between the participants' performance with the target two–step dependencies in Experiments 3 and 4 may have been caused by the presence of the successful one–step familiarity manipulation reducing outcomes in the latter case rather than by the difference in linguistic structure and the use of transfer in the former. Thus, the contrast between the results of the two experiments does not provide any evidence that the participants were sensitive to the difference between subject– and object–verb agreement. However future experimental research is needed to investigate the hypothesis.

10.4 Chapter Summary

The results of a series of simulations were consistent with the experimental findings and also extended the results by developing hypotheses for future testing. Simulations using a longer learning phase suggested that two–step dependencies can be acquired implicitly without transfer, but that longer ones cannot be regardless of the amount of input available. In addition the simulations clarified what it was possible to learn from the statistics of the input. Therefore they indicated that an alternative explanation of the participants' failure to acquire the bigrams was required for Experiment 5 but not for Experiments 2 or 3.

CHAPTER 11

General Discussion

11.1 Results Summary

This section synthesises the results of the experiments and simulations that make up this thesis with regard to the three research questions originally developed in Section 2.8, and reproduced here as (23).

(23) a) Is it possible to exclude explicit knowledge in order to obtain a valid measure of implicit knowledge?
b) Is there evidence of the acquisition of implicit knowledge in second language acquisition?
c) Does dependency length limit the acquisition of implicit knowledge of a second language?

11.1.1 Is it Possible to Exclude Explicit Knowledge in Order to Obtain a Valid Measure of Implicit Knowledge?

In the current experiments, implicit and explicit knowledge were distinguished according to three criteria. For performance to be classified as the result of explicit knowledge participants had to respond at above–chance levels in the correction test. Their performance also had to be indistinguishable from the controls in the timed judgement test but significantly more accurate than the controls throughout the untimed test. In contrast, performance was classified as implicit in origin if the participants were at chance in the correction test, and more sensitive than the controls in block one of the timed judgement test but not in block two. (See Section 3.3 for more details on how these criteria were developed.)

The procedure was based on the assumptions that the three criteria would all distinguish between the same two types of performance, and that those would be the result of implicit and explicit knowledge respectively. The second assumption depended to some extent on the first. If the criteria differed in the way in which they classified performance, it would be unclear which (if any) was actually differentiating between implicit and explicit knowledge and there could be little evidence of the former. Therefore this section discusses whether the criteria classified performance consistently.

In Experiments 1, 3 and 5 the rule detectors' judgement test performance met all the criteria for classification as being explicit in origin. By definition, the group was above chance in the correction test. They were not sensitive to grammaticality at any point in the timed test. However they were sensitive to grammaticality throughout the untimed test and without any noticeable decrease between blocks. The corresponding group in Experiment 2 demonstrated a marginal effect of grammaticality in the timed test that became significant in the untimed version. The group did perform marginally above the controls under time pressure, but the constraint nevertheless appeared to reduce their accuracy. Consistent with the use of explicit knowledge, there were no significant effects involving block in either test, suggesting that the incorrect input did not adversely affect performance. Finally in Experiment 4 the rule detectors did not demonstrate a significant effect of grammaticality in the timed test. However there was only a marginal one in the untimed test. There were no effects involving block (confirming that performance did not deteriorate following the introduction of incorrect input). Across all the experiments the rule detectors (who by definition met the first criterion of above chance performance in the correction test) were more sensitive to grammaticality in the untimed than in the timed test, and their performance was maintained in block two. Thus, the rule detectors' behaviour was always consistent with the use of explicit knowledge.

There were only sufficient familiarity detectors to permit statistical analyses in Experiment 1; those in the other experiments will not be discussed here. By definition their correction test performance suggested that the participants had explicit knowledge of the frequent bigrams, but examination of their judgement test performance was less conclusive. The analyses indicated that the group was significantly sensitive to familiarity in the timed test as well as in the untimed test, and the effect size was no smaller in the former than in the latter ($\eta_p^2 = 0.742$ and 0.678 respectively). Clearly the second criterion for explicit knowledge was not met as the time constraints did not produce an observable reduction in performance. Finally however, as their accuracy did not decline between blocks one and two of either test, their performance was consistent with the third criterion. In summary, the three criteria did not unanimously classify the familiarity detectors' performance in Experiment 1 as consistent with either implicit or explicit knowledge.

Unlike the rule and familiarity detectors, the non-detectors were assumed to be relying on implicit knowledge. They met the first criterion for this classification by definition as they were at chance in the correction test. Those in Experiments 2 and 5 were not significantly sensitive to either grammaticality or familiarity at any point in the judgement tests either, and can best be described as having neither any relevant implicit knowledge nor any relevant explicit knowledge. Therefore they are not included in this discussion. In Experiment 1 on the other hand the non-detectors were sensitive to familiarity in the timed test, thereby meeting the second criterion. However contrary to the third criterion they remained so in block two. The non-detectors in Experiment 3 were sensitive to grammaticality in block one of the timed test but not thereafter, fitting both remaining criteria for implicit knowledge. Finally, the performance of the non-detectors in Experiment 4 also met the criteria. They were sensitive to familiarity in block one of the timed test but not in block two. With the exception of Experiment 1 therefore, the criteria reliably classified the non-detectors' performance as consistent with implicit knowledge.

In all the experiments the rule detectors' performance met the three criteria for explicit knowledge, while the performance of the non-detectors in the language experiments fitted the three criteria for implicit knowledge. However the performance of the familiarity detectors in Experiment 1 was less clear–cut, as was that of the non-detectors in the same experiment. Further evidence was provided by the language background analysis, which uncovered an additional difference between the two types: only the development of explicit knowledge benefited from prior foreign language experience (see Section 9.2). Therefore we conclude that the proposed criteria did distinguish between two different types of performance.

Although the criteria differentiated between two types of performance the divide may not have been between the use of implicit and explicit knowledge. Such an interpretation relies on an additional assumption. Rather than directly tapping the presence or absence of conscious awareness, the criteria were based on the proposal that consciousness provides control over behaviour (Cleeremans & Jiménez, 2002). If this is incorrect, then the two types of performance that were identified would be unlikely to reflect the use of implicit and explicit knowledge respectively. However, even under this weaker interpretation the experiments still provided evidence of two separate types of performance and therefore the findings would remain of theoretical interest.

11.1.2 *Is There Evidence of the Acquisition of Implicit Knowledge in Second Language Acquisition?*

The non-detectors in Experiment 3 demonstrated implicit knowledge of the target subject–verb agreement according to the three criteria. Similarly, those in Experiment 4 performed as if they had implicit knowledge of the frequent pairs of adjacent words. However there was no evidence of any relevant implicit knowledge in the other language learning experiments. (The non-detectors in Experiments 2 and 5 were indistinguishable from the controls throughout.) Nevertheless the results of Experiments 3 and 4 suggest that learners can acquire implicit knowledge of second language structures, albeit under limited circumstances.

The finding that second language learners can acquire implicit knowledge is consistent with those second language acquisition studies that contrasted either learning under incidental conditions or the development of implicit knowledge with chance performance. See Section 2.3 for a more complete review, but to highlight some examples the implicit groups trained by Bitan and Karni (2004) and Michas and Berry (1994) acquired grapheme–phoneme conversion rules. In a different vein, those tested by Erlam (2006) demonstrated implicit knowledge of a series of English structures. Overall therefore the current results are consistent with previous findings that adult second language acquisition can include the development of implicit knowledge (Seliger, 1979; Williams, 2005; Erlam, 2006; Hulstijn & Hulstijn, 1984). Note however that performance based on explicit knowledge tended to be more accurate than that resulting from implicit knowledge as has also been found in many previous studies (see Norris & Ortega, 2000, for a review). Importantly however the current findings were not based on verbal report which has a tendency to overestimate implicit knowledge, but rather on three independent criteria.

11.1.3 Does Dependency Length Limit the Acquisition of Implicit Knowledge of a Second Language?

As discussed above in Sections 2.4 and 2.5, the acquisition of implicit knowledge of a particular artificial grammar or sequence is strongly influenced by dependency length in the absence of specific attentional manipulations. When the target structure is a contingency between adjacent sequential items learning is usually successful, but when additional elements intervene performance is reduced (Cleeremans & McClelland, 1991). Despite differences in procedure, similar results have been found in statistical learning experiments, where participants learn to segment a continuous speech stream according to the transitional probabilities between syllables (Newport & Aslin, 2004). This section asks whether dependency length has a similar influence in the development of implicit knowledge in adult second language acquisition.

A contingency is a rule determining the form of one element in a sequence based on that of another. The length is calculated according to the amount of material intervening between the two. This intervening material is not of predictive value. In a one–step dependency the two elements are adjacent so there is only a single step. In a two–step version on the other hand there is one intervening element (and two steps, from the first to the intervening one and from the intervening one to the second). In artificial grammar learning experiments each element is considered to be a single letter, so that a one–step dependency is a contingency between adjacent letters. In the statistical learning literature on the other hand, each element is normally considered to be a syllable (Peña et al., 2002). Thus, a one–step dependency is a rule connecting adjacent syllable. As the input in the Experiments 2 — 5 was segmented into individual words, we assumed that participants were likely to segment it as such. Therefore we considered an element to be equal to a word, and measured the dependency length accordingly (cf. Robinson, 2005). However if participants did segment the input into smaller units, the contingency lengths would have been longer than those specified.

In Experiment 4 the non-detectors demonstrated implicit knowledge of the frequent bigrams (a regularity that can be represented as a one-step dependency — see Section 3.2). Their counterparts in Experiment 3 had acquired implicit knowledge of a two–step dependency representing subject–verb agreement. However those in Experiment 5 did not learn the latter construction when it was a three–step dependency. This pattern of results is consistent with the interpretation that implicit knowledge of one– and two–step contingencies may be acquired but that longer ones may not be. Thus it suggests that, as in the other experimental paradigms, the length of a dependency has a substantial influence on whether implicit knowledge is acquired. (Note however that neither the one– nor the two–step dependencies were acquired consistently. There was no evidence that the non-detectors in Experiment 5 had any knowledge of the frequent bigrams, and nor did the group in Experiment 2 acquire the two–step subject–verb agreement. Unsurprisingly therefore additional factors as well as dependency length must influence performance.)

One aspect of identifying the lengths of dependency for which implicit knowledge can be acquired is demonstrating that certain structures are not learnt. This project has shown that a three–step dependency is not learnt implicitly within the confines of a relatively short experiment. Clearly the conclusions would have been strengthened had the learners received more input. As discussed in Section 3.4, increasing the exposure given to participants would have been problematic because additional training would have removed two of the distinguishing features of implicit knowledge. Therefore the effect of lengthening the exposure phase was addressed in Chapter 10 using simulations.

Following sixty thousand learning trials the connectionist network acquired sensitivity to the majority of the one–step regularities in the data, with the exception of the Persian bigrams (see Section 10.3 for a discussion of this result). It also acquired the two–step dependencies, regardless of whether they represented subject–verb agreement (as in Experiment 3) or object–verb agreement (as in Experiment 4). However there was no suggestion that the network became sensitive to the three–step dependency in Simulation 5, despite the increased amount of input. Thus, the simulations were consistent with the interpretation that contingency length influences performance and specifically that statistical learning can succeed with one– and two–step dependencies but not with longer structures.

The network's success modelling human performance with the shorter learning phase raises the possibility that the participants were also using a frequency–tallying approach. Thus, the extended simulations provided a prediction of human performance under parallel situations. In particular, they suggested that statistical learning mechanisms would be unlikely to allow a learner to acquire a three–step dependency. If participants in future experiments do acquire the structure therefore, other types of learning are likely to be contributing to their performance. On the other hand if they do not, the findings would remain consistent with a role for statistical learning in adult second language acquisition.

Overall, the findings are consistent with experiments in other paradigms. As reported above in Section 2.6, adults and infants in statistical learning experiments quickly acquire sensitivity to one–step dependencies (Saffran, Aslin, & Newport, 1996). Their performance with two–step dependencies is also successful under certain circumstances (Gómez, 2002; Peña et al., 2002), but not in all cases (Newport & Aslin, 2004). Performance is similar with non-linguistic input in the guise of both artificial grammars and sequences of locations in the SRT task. Participants acquire implicit knowledge of one–step dependencies (Nissen & Bullemer, 1987; A. S. Reber, 1989). They can also learn trigrams, (Frensch et al., 1994; Reed & Johnson, 1994a; Schvaneveldt & Gomez, 1998). However they do not acquire longer regularities as confirmed by the current Experiment 1 and by Cleeremans and McClelland (1991), Johnstone and Shanks (2001), Mathews et al. (1989) and St. John and Shanks (1997). In all cases therefore, one–step contingencies are acquired and trigrams and two–step dependencies are under limited circumstances, but there is no evidence of implicit knowledge of longer structures in the absence of specific attentional manipulations. The parallel between the different experimental paradigms raises interesting questions as to whether similar learning mechanisms or processes are employed in each case.

In summary, the series of experiments and simulations in this thesis suggested that implicit knowledge can be acquired of one– and two–step dependencies but not of three–step ones.

11.2 Generalisability of the Results

This section examines the extent to which the current findings reflect second language acquisition as it occurs in other experiments and outside of the laboratory. It considers the role of the input statistics in language learning and whether language knowledge is abstract or surface–based. It also examines claims that attention is a prerequisite for implicit knowledge to be acquired. Finally, it discusses the similarities and differences between the current target structures and a full natural language.

11.2.1 *Frequency in Language Acquisition*

The current experiments investigated a learning system that is based on the statistics of the input (cf. N. C. Ellis, 2005). The non-detectors' performance in Experiment 4 provides a clear example of the statistical origins of their behaviour. The group had acquired the frequent bigrams, but once enough incorrect input had been provided that the familiar and unfamiliar items could no longer be distinguished by the ACS values, test performance returned to chance levels. The similarity between the results of the experiments and their corresponding simulations is also consistent with a statistically driven process, as such models are driven largely by statistics. This section asks whether such a learning system is typical of language acquisition and use in general and of adult SLA in particular.

Frequency plays a major role in language processing (see N. C. Ellis, 2002a, 2002b, 2006, for a more complete review). However to summarise, frequent words are pronounced more quickly than their less frequent counterparts (Balota & Chumbley, 1984). Phonological reduction occurs to a greater extent in frequent than in infrequent contexts (Bybee, 2002), and the frequency with which words are used in different constructions helps listeners resolve ambiguous utterances (Seidenberg & MacDonald, 1999; Spivey & Tanenhaus, 1998). Infants acquiring their first language are also highly sensitive to statistical information (see Gómez & Gerken, 2000, for a review). Thus, frequency effects are found both in first language acquisition and in native language processing.

A number of studies have shown that frequency also affects the performance of second language learners. Firstly like infants, adults are sensitive to the statistical distribution of nonsense syllables in the speech stream (e.g. Gómez, 2002; Saffran, Newport, & Aslin, 1996). Secondly, learners of Russian and Norwegian are able to produce the past tense of nonce verbs in frequent verb classes more accurately than those in less frequent verb classes (Tkachenko, 2007, September). Thirdly, frequency influences the order of acquisition of English grammatical morphemes in SLA (Goldschneider & DeKeyser, 2001). Finally, N. C. Ellis and Schmidt (1998) reported that participants produced plural forms of frequent nouns in a miniature artificial language more accurately than those of infrequent nouns. Currently there is little direct evidence on the role of frequency in the acquisition of *implicit* knowledge in a second language, but it is likely to influence performance.

One study appears to contradict the claim that input statistics influence adult second language acquisition. Robinson (2005) reported that experienced language learners were no more likely to accept Samoan sentences with higher Associative Chunk Strength (ACS) as grammatical than those items with lower ACS (see Section 4.1.2 for details of the calculation). However these results must be treated with caution. Unlike Johnstone and Shanks (2001), Robinson did not appear to include the onset and offset bigrams (each including the first or last word in the utterance and a boundary marker) in his calculation of ACS. Additionally, the example calculation he provided appeared to omit one of the remaining bigrams. For these reasons it is not clear whether his distinction between high and low ACS items was valid, and therefore whether his conclusion that frequency does not affect SLA was justified.

On balance, the evidence suggests that frequency is an important factor both in first and in second language acquisition. These effects are found both for classroom learners and in controlled experimental situations. Thus, the statistical learning process investigated in the current experiments may also be employed in second language acquisition in more naturalistic settings. Nevertheless this statement must not be interpreted as a claim that language acquisition is *purely* statistically–driven. Indeed, it is understood to interact with many other factors (Gass & Mackey, 2002; Hulstijn, 2002b).

11.2.2 Abstract versus Surface Knowledge

A long–standing debate in artificial grammar learning research is whether it is possible to acquire implicit knowledge of rules based on abstract categories as well as those that depend on surface items (see for example Altmann et al., 1995; A. S. Reber, 1969; Shanks et al., 1997). This question is also relevant to language acquisition. Archetypal linguistic rules (including noun–verb agreement as utilised in the current experiments) are abstract. They rely on categories such as *subject*, *object* and *verb* that must be abstracted from the input. Thus, abstract knowledge may be more common in second language acquisition than in other types of learning. A surface regularity on the other hand would relate to specific lexical items or morphemes. This section discusses whether the implicit knowledge that was acquired during the current experiments related to surface or abstract categories.

Abstractness was not the focus of the current work, so the materials did not allow us to distinguish between abstract and surface–based knowledge of the target structures. Every subject (and object) noun in the learning phase materials was paired with every finite verb. Correct classification of a judgement test item could be based either on an abstract contingency such as *singular subject — x — singular verb*, or on a series of surface patterns such as *mehmun — x — keshe* and *yateem — x — keshe*. Unlike the target structures however, the frequent bigrams could only be detected using surface–based knowledge because they depended on the individual lexical items. In Chapter 7 we argued that the non-detectors in Experiment 4 had acquired implicit knowledge of the repeated word pairs. Therefore the group cannot have been using abstract knowledge. The remainder of this section considers whether this finding is consistent with performance in SLA outside of the laboratory.

As already stated, prototypical linguistic rules apply to abstract categories. Nevertheless many other linguistic regularities are based on individual lexical items. Estimates suggest that up to half of native language production consists of formulaic sequences of surface items (Schmitt & Carter, 2004). These are used in reading comprehension (Underwood, Schmitt, & Galpin, 2004). Such sequences are also common in learner speech, particularly at lower proficiency levels (Wray, 2004). Turning to acquisition, Gómez and Lakusta (2004) discovered that adult participants in a statistical learning experiment could only acquire regularities based on abstract categories when additional cues were available. Therefore the majority of learners' performance in other experiments without those cues is likely to have been the result of surface–based knowledge. In summary, regularities between lexical items contribute to production, comprehension and second language acquisition. Therefore the non-detectors' knowledge of the frequent bigrams was consistent with aspects of language acquisition and use outside of the laboratory.

11.2.3 Attention

There is no evidence that the trained participants acquired implicit knowledge of the frequent bigrams in Experiment 5. As discussed above in Section 8.3, this was unexpected because the ACS contrast between the familiar and unfamiliar stimuli was larger than in Experiment 4, where implicit knowledge of the corresponding pattern was acquired. In addition, Section 10.2.5 reported a simulation in which a connectionist network became sensitive to familiarity when modelling Experiment 5, confirming that it could theoretically be acquired from the statistics of the input after an equally–sized learning phase. Thus, a factor other than bigram frequency must have influenced learning outcomes.

A tentative explanation for this result was offered based on attention. Specifically, the two–part proposal was that the acquisition of implicit knowledge requires attention, and that the participants focused their attention on the words that underwent morphological variation. These two factors will be discussed in turn.

The development of implicit knowledge in SLA is assumed to depend on attention (Schmidt, 1993; Tomlin & Villa, 1994; Williams, 1999), a claim for which there is limited experimental evidence. In a study carried out by Erlam, Ellis, and Loewen (2007, September), learners of English as a second language did not acquire a structure from which their attention was diverted during input flooding. Hulstijn and Hulstijn (1984) also discovered that attention must also be paid to the relevant aspect of the grammar rather than to the meaning for relevant implicit knowledge to be detected.

The dual–task procedure is often used to investigate the importance of attention in SRT experiments. Two groups of participants perform the primary SRT task, while one of them concurrently completes a secondary task to deplete their attentional resources. The secondary task sometimes reduces or eliminates the development of implicit knowledge in the serial response time paradigm (Carr & Curran, 1994; Cohen et al., 1990; Nissen & Bullemer, 1987; Shanks et al., 2005), although alternative explanations of such results have been offered (Jiménez & Vazquez, 2005; Schmidtke & Heuer, 1997). Similar experiments have demonstrated that divided attention can reduce performance in the statistical learning paradigm (Toro et al., 2005). Using a different technique, Jiménez and Méndez (1999) and Jiang and Chun (2001) discovered that implicit knowledge is only acquired when selective attention is paid to the relevant aspect of the input. Overall, the evidence reported in this section implies that the development of implicit knowledge depends on the availability of attention.

We now consider whether the participants were paying sufficient attention to the relevant aspects of the input to acquire the frequent bigrams in Experiment 5. It is likely that they focused more attention on the words that underwent morphological variation than on those that did not. In Experiment 5 the target structure was subject–auxiliary agreement. Only the subject noun and the auxiliary verb varied in number while the novel bigrams employed in the unfamiliar items consisted of the object and the lexical verb. Therefore, neither of the words in the frequent bigrams underwent morpological variation, and therefore insufficient attention was paid to the bigrams for them to be acquired. In Experiment 4 on the other hand, the target structure was object–verb agreement, so the object noun underwent morphological variation. This word also formed half of the frequent bigrams. Therefore the participants were likely to have paid more attention to those bigrams, which allowed them to acquire the target structure. In summary therefore, we claim that the non-detectors did not acquire the frequent bigrams in Experiment 5 because they were not paying sufficient attention to the relevant section of the stimuli. Nevertheless, the account must remain tentative in the absence of further research.

11.2.4 Target Structures

The modified languages employed in Experiments 2 — 5 were miniature human languages, but they had been simplified to a large degree. This section considers the extent to which they provided a valid model of natural language. In particular it focuses on recursion and hierarchical structure, both distinguishing features of natural language (M. D. Hauser, Chomsky, & Fitch, 2002).

The experimental target structures were subject and object–verb agreement, instantiated either as two– or three–step dependencies. Both are found in natural languages. (The former is common, while the latter is found in languages including Basque and Hungarian.) As each item used in an experiment followed an identical structure with only a single instance of the target structure, there was no evidence of recursion. Similarly, there was no evidence that the target structure was hierarchical rather than sequential. For example, the modified Persian subject–verb agreement could be described as a dependency between the first and the third words, rather than between the subject and the verb. Thus, it is important to note that the participants may have acquired a sequential two–step structure rather than a hierarchical or grammatical–role based dependency. On the other hand if the number of intervening words had varied between items, a sequential interpretation of the structure would not have been tenable. It is important to consider whether participants engage the same learning modes when learning hierarchical and non-hierarchical structures, in order to ascertain whether the current experiments provided a valid model of language acquisition.

There is evidence that the brain may process hierarchical and sequential structures differently. Friederici, Bahlmann, Heim, Schubotz, and Anwander (2006) trained adult learners on either a hierarchically structured artificial language or a sequential version, before administering a grammaticality judgement test during which functional Magnetic Resonance Imaging (fMRI) was used to determine the areas of the brain contributing to performance. The participants exposed to the hierarchical language later engaged Broca's area (known for linguistic processing) when reading incorrect stimuli, but those trained on a sequential version did not. Thus, Friederici et al. concluded that language learning modes are only engaged when hierarchical structure is encountered. If this was accurate, the current results obtained using sequential target structures, would not be generalisable to language learning outside of the laboratory.

There are two alternative interpretations of the data provided by Friederici et al. (2006). Firstly, the hierarchical version of the language included long–distance dependencies, whereas the sequential structure did not. Rather than linguistic structure therefore, these long–distance dependencies may have triggered the involvement of Broca's area and presumably a language learning mode. Secondly, the learners were presented with both grammatical and ungrammatical stimuli during the learning phase, and they were instructed to identify the target rule by trial–and–error. Therefore it is highly likely that the participants learnt explicitly and the results may not generalise to the acquisition of implicit knowledge.

In contrast to Friederici et al. (2006), Vries, Barth, Knecht, Zwitserlood, and Floeel (2008, September) conducted a transitional artificial grammar learning experiment using two groups of trained participants. The trained group received electrical stimulation to Broca's area during the learning phase, which was predicted to improve performance only if Broca's area was engaged in the task. The control participants did not receive this stimulation, although they were exposed to the same input. The group that received stimulation outperformed the controls in a subsequent grammaticality judgement test. Thus the authors concluded that the region contributed to performance, although the target structure was neither linguistic nor hierarchical. Supporting evidence was also provided by Christiansen and Ellefson (2002), who found that aphasic patients with linguistic difficulties were also impaired in tests of (non-linguistic) artificial grammar acquisition. Overall therefore, the importance of hierarchical structure and recursion for the engagement of linguistic processing modes or mechanisms is still an open empirical question.

In summary, although they are based on real languages, the experimental materials do differ from natural language in some respects. Nevertheless, the extent to which these differences influence processing and acquisition and therefore the importance of using highly naturalistic input in laboratory SLA studies is still an open empirical question.

11.2.5 Section Summary

Overall the type of learning detected in these experiments is likely to contribute to SLA outside of the laboratory. Frequency effects, surface knowledge and the importance of attention are all features of learning in other situations. However, not all SLA can be explained in this way. As discussed above in Chapter 2, form–meaning relations and complex syntactic structures are unlikely to be acquired from the statistical distribution of the input forms alone. Rather, deliberate and explicit hypothesis testing usually contributes to adult second language acquisition (DeKeyser, 1995), while the learners may also have access to universal grammar (Kanno, 1997; Schwartz & Sprouse, 1996).

11.3 Comparability with First Language Acquisition

The work in this thesis was focused on the acquisition of implicit knowledge in adult second language acquisition. Nevertheless it is also worth considering the extent to which the processes and mechanisms investigated also function in first language acquisition (FLA), or alternatively whether they depend on prior linguistic knowledge.

As discussed in Section 2.6, there are many correspondences between research in the statistical learning paradigm in which adults and infants learn to segment speech streams based on the statistical distribution of the syllables (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and the current series of experiments. In statistical learning experiments the statistics provide the only evidence of the correct segmentation. In the current experiments on the other hand, the participants were given additional sources of information. For example, rather than acquiring a dependency between the forms of the plural markers on the relevant noun and the verb, they may have used form–meaning relations to compute the correct form independently for each word. Nevertheless the participants' performance in these experiments was successfully modelled by a series of connectionist simulations, a finding consistent with a substantial influence of the input statistics on their resulting knowledge. In addition, dependency length appears to affect experimental performance in both paradigms. Performance with one–step dependencies is robust in statistical learning experiments (Gómez & Gerken, 2000; Saffran, 2003). However Gómez (2002) and Peña et al. (2002) reported that two–step ones were only acquired under limited circumstances, while Newport and Aslin (2004) failed to find any evidence that they were acquired at all. The participants in the experiments in this thesis acquired implicit knowledge of one– and two–step dependencies under limited circumstances, but they did not acquire the three–step one in Experiment 5.

Although the thesis was not designed to test the possibility, these similarities suggest that the adult second language learners we tested were using processes or mechanisms that are also available to the infants who participated in many of the statistical learning experiments (Saffran, Aslin, & Newport, 1996). Nevertheless, although the learning mode is probably available to infants, it does not automatically follow either that they use it in FLA or that it does not depend on prior linguistic knowledge. To the best of our knowledge the youngest infants tested in these experiments were seven months old (Thiessen & Saffran, 2003). Although they were still in the initial stages, they had already begun the process of language acquisition.

In summary, the findings are consistent with claims that the same mechanism contributes to FLA. Nevertheless just as with SLA, statistical learning mechanisms focusing on the co-occurrence of elements in the input may be able to acquire agreement relations but they cannot account for the entire acquisition process. Form–meaning relations, discourse requirements and complex syntactic relations are all likely to be acquired by different means.

11.4 Remaining Issues

This section discusses three further methodological issues: the power of the experiments, discrepancies in the number of experimental items and the validity of manipulating familiarity in addition to grammaticality.

11.4.1 Experimental Power

There were eighteen trained participants in each of Experiments 4 and 5. In the former case seven of those were non-detectors, and in the latter eight were. Only the non-detectors' performance was analysed in order to detect implicit knowledge. Therefore the conclusions drawn in these experiments were only based on seven and eight trained participants respectively. This has particularly important consequences for the null effects.

There was no evidence that the trained participants had acquired implicit knowledge of object–verb agreement in Experiment 4, of subject–verb agreement in Experiment 5, or of the frequent bigrams in Experiment 5. In all cases it is important to consider whether there really was no relevant implicit knowledge, or whether the results were only non-significant because of a lack of power. In all cases we consider performance in block one of the timed test, which is where implicit knowledge was assumed to be utilised.

In all three cases, the non-detectors were less sensitive to the structure than were the control participants. Therefore it was not possible to conduct a power analysis to identify the ideal sample size in order to detect significant implicit knowledge. This calculation is based on the effect size. As the effect obtained was in the opposite direction to that predicted, no measure of the size of the predicted effect was available and the calculation could not be carried out. However, as the results were qualitatively rather than quantitatively different from the prediction, additional participants would have been unlikely to alter the outcome. Nevertheless had more time been available during the project, we would have included more participants in these experiments to confirm the findings.

11.4.2 Quantity of Experimental Items

There were different numbers of learning and test items in the Persian materials (used for Experiments 2 and 3) and the Basque materials (used for Experiments 4 and 5). Specifically, there were forty learning items in the Persian materials, and twenty–four in the Basque materials. Each Persian judgement test contained forty items, while the Basque judgement tests had forty–eight. Although both correction tests contained twelve items, in the Persian experiments only four of those items were experimental and the remaining eight were fillers. In the Basque experiments the stimuli were equally split, with six experimental items and six fillers.

The origin of the majority of these differences was as follows. Experiments 2 and 3 were the first to be administered, and therefore the Persian materials the first to be designed. When the biconditional grammar experiment was planned the materials were adapted from an experiment by Johnstone and Shanks (2001) that had a different number of items. The quantity was then held constant in the Basque materials. With hindsight however, as the primary comparison was between the linguistic experiments rather than between the linguistic and artificial grammar experiments, this was not the ideal strategy. Instead, it would have been advisable to match the number of items across the two sets of linguistic materials.

It is hoped that the differences in the number of stimuli did not have an undue effect on performance. Experiment 5, the results of which suggested that implicit knowledge of three–step dependencies is not acquired, contained the smaller number of learning items. However Simulation 5b indicated that extending the exposure phase did not improve network performance. If the parallel between experimental and simulation results were to be maintained, the same may also be true of human performance. Nevertheless, increasing the number of unique learning items may have had a different effect than simply increasing the number of tokens through repetitions.

11.4.3 The Validity of the Familiarity Manipulation

It is common practice in artificial grammar learning research to manipulate familiarity (sometimes termed *ACS*) orthogonally to grammaticality in judgement test items (e.g. Channon et al., 2002; Johnstone & Shanks, 2001; Robinson, 2005). This custom stems from concerns that participants classify strings resembling the training strings as correct. If the grammatical items were more similar to the training strings than were the ungrammatical items, such behaviour could otherwise masquerade as knowledge of the target rule. Manipulating the two orthogonally rules out such a possibility. For the same reason, this approach was taken in the current thesis.

The results of Experiment 4 and its simulations may cast doubt on the validity of manipulating familiarity in this way. As already reported, the participants appeared to have acquired implicit knowledge of the frequent bigrams. However there was no evidence that they had any implicit knowledge of the target two–step object–verb agreement. In contrast, in Experiment 3 where familiarity was not successfully manipulated, the participants did acquire implicit knowledge of the two–step subject–verb agreement. Therefore the length alone could not account for the pattern of results. Instead it was proposed that the participants could not acquire the target structure in Experiment 4 within the short exposure period because they were focusing on the bigrams instead. This hypothesis was supported by the results of Simulations 4a and 4b, as the network was only able to acquire object–verb agreement once the exposure phase had been extended. There is also independent evidence that longer dependencies can only be acquired after shorter ones, at least in the serial response time task (Cleeremans & McClelland, 1991; Schvaneveldt & Gomez, 1998).

If introducing a shorter contingency delays the acquisition of a longer one as suggested, then the use of familiarity to control for confounds with a longer target structure is not valid. The distraction provided by the shorter pattern could prevent the participants acquiring a longer dependency that they would otherwise have learnt. As such, the technique would underestimate the extent to which implicit knowledge of long–distance dependencies can be acquired. Further research should investigate this issue, confirming the hypothesis drawn from Simulation 4a and 4b, for example by replicating the corresponding experiment without the familiarity manipulation.

11.5 Chapter Summary

In summary, the current experiments distinguished between two types of performance that fitted the expected characteristics of implicit and explicit knowledge respectively. They provided evidence that participants could acquire implicit knowledge of one– and two–step dependencies but not of three–step ones. Interestingly this limit is consistent with results from artificial grammar learning, the serial response task and statistical language learning experiments, despite the major differences in target structures and in experimental procedures.

Nevertheless some methodological concerns remain, relating to the power of the experiments that produced the crucial null results, differences in the number of items between the Persian and Basque experiments and the use of the familiarity manipulation. As is always the case in research into implicit learning and implicit knowledge, there is also a question about whether the two types of performance identified were actually the result of implicit and explicit knowledge, or whether they should be characterised differently. For these reasons it is important to replicate these findings in future research.

The type of implicit knowledge investigated in this thesis depended strongly on the statistics of the input, and was also shown to require attention. It was not possible to ascertain whether the participants acquired any abstract categories or hierarchical structures, but they did learn at least one pattern based on the surface forms that may have been represented sequentially. Although second language acquisition is not limited to learning of this type, these features are nevertheless characteristic of the process as it occurs outside of the laboratory. Thus, the experiments provided valid information on the acquisition of implicit knowledge in adult second language acquisition.

CHAPTER 12

Conclusion

12.1 Summary of Empirical Findings

This investigation into the role of implicit knowledge in adult second language acquisition contributed to both theoretical and to methodological issues. Firstly it presented new empirical data to identify the types of structure that are amenable to the acquisition of implicit knowledge. Secondly it developed a new method to distinguish implicit from explicit knowledge based on three independent criteria. These two aspects of the work will be discussed in turn.

Section 2.1 presented evidence that implicit knowledge is acquired continuously (Ferreira & Bock, 2006; Frensch & Miner, 1994). Nevertheless, previous evidence suggests that performance based on implicit knowledge alone is unlikely to lead to native–like or even to advanced proficiency levels (Norris & Ortega, 2000). This thesis investigated one possible explanation of this seeming contradiction: whether the acquisition of accurate implicit knowledge is restricted to certain types of target structure.

Dependency length may be a factor that influences whether implicit knowledge of a structure can be learnt. Participants acquire contingencies between adjacent items in artificial grammar learning, serial response time and statistical learning experiments (A. S. Reber, 1967; Nissen & Bullemer, 1987; Saffran, Aslin, & Newport, 1996). However performance is reduced in all three paradigms when the target structure is longer (Cleeremans & McClelland, 1991; Newport & Aslin, 2004; Perruchet & Pacteau, 1990). The first experiment in this thesis replicated previous findings that longer structures are not acquired in the AGL context. The remaining four then investigated whether contingency length also affects the acquisition of noun–verb agreement in a novel second language.

The experimental results suggested that learners could sometimes acquire implicit knowledge of one– and two–step dependencies in a second language, but that three–step contingencies were not acquired. Consistent data were also generated in a series of connectionist simulations. Thus, the pattern of results was interpreted as evidence that dependency length influences whether implicit knowledge of a given structure can be learnt. Nevertheless, it was not the only factor that influenced the results. Implicit knowledge was only acquired in one of the two learning tasks employed, while either transfer from the first language or the presence of additional regularities in the input also affected performance. Section 12.2.1 will describe future experiments that could clarify the role of these factors.

Previous empirical attempts to distinguish implicit from explicit knowledge have proved both controversial and inconclusive. It is widely recognised that no test can detect the full extent of either type, and nor can it completely exclude use of the other (Jacoby, 1991; Shanks & St. John, 1994). Thus, whenever performance is significantly above chance on a test designed to detect implicit knowledge, the data may be the result of the intrusion of explicit knowledge. In contrast, the current experiments employed three independent criteria based on performance in three different tests. In the language–based experiments the criteria tended to classify performance unanimously as the result of either implicit or explicit knowledge. Thus, this approach strengthened the conclusion that the adult learners had acquired implicit knowledge of a second language.

12.2 Recommendations for Future Research

This section makes two types of recommendations. Firstly, it proposes specific experiments that would strengthen and extend the findings of the current study. Secondly, it recommends general directions for future research into the acquisition of implicit knowledge in adult second language acquisition.

12.2.1 Specific Issues Arising from this Work

The research in this thesis was consistent with the interpretation that learners can acquire implicit knowledge of one– and two–step dependencies, but not of longer contingencies. Nevertheless alternative interpretations of the data remain, and additional experiments could clarify the conclusions.

The finding that implicit knowledge of subject–verb agreement could be acquired as a two–step dependency came from Experiment 3, which used the Persian materials. Experiment 5, which demonstrated that a similar structure was not acquired as a three–step dependency, utilised the Basque set. Unfortunately there were differences between the two sets of materials. Crucially, there were more learning items in the Persian than in the Basque one. In addition the lexical verb in the Persian sentences was finite, whereas in the Basque items it was a past participle. The first recommendation is to test the acquisition of implicit knowledge of two–step subject–verb agreement using the Basque materials. If the results parallel Experiment 3, the current conclusions would be strengthened. Otherwise, it would be necessary to investigate other potential sources of the present effects.

The non-detectors in Experiment 3 were able to acquire implicit knowledge of a two–step dependency corresponding to subject–verb agreement, but those in Experiment 4 did not acquire a structure of the same length representing object–verb agreement. Two explanations were offered for this pattern of results. Either transfer from the first language allowed subject–verb but not object–verb agreement to be acquired, or else the presence of an additional one–step regularity in Experiment 4 slowed or reduced learning of the longer structure relative to Experiment 3. The first explanation would imply that the learners were sensitive to the underlying linguistic structure in order to distinguish between subject–verb and object–verb agreement. However, the connectionist network's performance supported the second interpretation. Nevertheless, conclusions about human abilities derived from simulations must remain tentative.

Two further experiments may be conducted to address the contribution of each of these factors. One would examine the acquisition of two–step subject–verb agreement when there is also a valid one–step contingency in the input data. If performance is reduced relative to the current Experiment 3, it would suggest that the presence of a shorter (and presumably simpler) pattern affects the acquisition of longer ones. The second experiment would investigate the acquisition of two–step object–verb agreement without any other regularity in the input. To the extent that performance is lower than that in Experiment 3, it would indicate that transfer aids the acquisition of implicit knowledge of two–step linguistic contingencies.

Conducting the three experiments proposed in this section would clarify the results of the current experiments, thereby strengthening the conclusions drawn. In particular, it would replace data from simulations with that provided by human participants.

12.2.2 General Directions for Future Research

The current project characterised structures as acquirable or not based on their position on one dimension: the length of the contingency. Future research should continue this approach, asking what characteristics a target structure must have in order to be acquired rather than focusing on individual structures. Such a strategy would provide a coherent body of literature within which individual studies could be clearly interpreted.

The target structures used in the current experiments were two– or three–step dependencies, and the length was held constant for every item within an experiment. Exposure to naturalistic input on the other hand would provide exemplars of the same structure at varying lengths. For example, a learner would encounter a mixture of utterances such as (24*a*) where subject–verb agreement is a one–step dependency, versions such as (24*b*) where it is a two–step dependency and the seven–step version in (24*c*). One important question therefore is whether implicit knowledge of a structure can be acquired from shorter exemplars and then applied in longer versions. Gómez (2008, September) reported initial research into this issue in the statistical learning paradigm. She claimed that twelve–month old infants could acquire a one–step dependency, and then apply their resulting knowledge to a two–step version that they would otherwise not acquire. Whether adult second language learners have a similar ability is still an open question.

(24) a) John walks home.
b) John always walks home.
c) John, the man I saw yesterday, always walks home.

Contingency length is not the only factor that influences whether implicit knowledge of a given second language structure is acquired. Therefore similar research is needed with different types of target structure. For example, form–form contingencies should be compared with form–meaning relations, and morphological rules with basic word order patterns. In addition, previous research into artificial grammar and statistical learning has indicated that learners can acquire abstract categories (Altmann et al., 1995; Gómez & Lakusta, 2004; A. S. Reber, 1969, 1989). This is also relevant to second language acquisition, as most linguistic rules and regularities apply to word classes rather than to the surface lexical items. Future research may investigate whether adult second language learners can acquire implicit knowledge of such patterns.

Finally, this thesis provided evidence that contingency length influences the acquisition of implicit knowledge of second language structures. Other researchers have previously suggested that a similar constraint applies to performance in three other experimental paradigms: artificial grammar learning, the serial response time task, and statistical learning (Mathews et al., 1989; Cleeremans & McClelland, 1991; Newport & Aslin, 2004). Thus, the current findings were consistent with the claim that similar learning processes are engaged in each case. Although this possibility has been addressed directly by Friederici et al. (2006) and Robinson (2005), the results are not yet conclusive. Therefore further research is needed, directly comparing abilities with each type of input across a wider range of target structures and learning situations.

Appendix A

Participant Information

Tables A.1 and A.2 present background information on the experimental participants, divided between experiments and between trained and control groups. The data include the participants' sex and age, the number of foreign languages they had learnt prior to the experiment, and their current ability in their best foreign language (self–reported on the scale in Chapter 4).

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
Total	16	36	36	18	18
Males	8	12	18	8	7
Age	22 (20 — 23)	21 (20 — 22)	21 (19 — 22)	21 (19 — 22)	20 (19 — 21)
No. langs.	2 (2 — 3)	2 (1 — 2)	2 (1 — 2)	2 (1 — 2)	2 (1 — 2)
Ability	3.75 (3 — 5)	3 (2 — 3)	3 (2 — 3)	3 (2 — 3)	3 (2 — 3)

Table A.1: Trained participant details. The table shows the median (and interquartile range in parentheses) of the participants' ages, number of foreign languages encountered, and current ability in their best foreign language.

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
Total	16	36	N/A	18	18
Males	8	18	N/A	10	8
Age	21 (21 — 22)	21 (20 — 23)	N/A	20 (19 — 21)	22 (21 — 24)
No. langs.	2.50 (2 — 3)	1 (1 — 2)	N/A	2 (2 — 2)	2 (2 — 3)
Ability	4.75 (4 — 5)	3 (2 — 3)	N/A	3 (3 — 4)	3 (3 — 4)

Table A.2: Control participant details. The table shows the median (and interquartile range in parentheses) of the participants' ages, number of foreign languages encountered, and current ability in their best foreign language. Experiment 3 used the same control participants as Experiment 2.

Language	Taught Learners	Naturalistic Learners	Total Learners
French	209	28	215
German	109	12	111
Spanish	63	22	73
Latin	22	0	22
Chinese	6	6	10
Arabic	6	5	9
Japanese	4	3	6
Other Romance	**16**	**8**	**23**
Celtic	**11**	**2**	**12**
Slavic	**5**	**7**	**10**
Other Germanic	**7**	**5**	**9**
Austronesian	**3**	**6**	**8**
Bantu	**1**	**6**	**7**
Other	**8**	**16**	**20**
TOTAL	**470**	**126**	**535**

Table A.3: Second languages learnt prior to the experiment, based on the information given in the questionnaire. Taught Learners said they had had explicit instruction (including at school or from teach–yourself books). Naturalistic learners were exposed to the language through contact with native speakers (for example on holiday, or from foreign friends in the UK). Some participants noted both explicit instruction and naturalistic exposure for the same language. Therefore the taught and naturalistic columns do not sum to give the total. Also note that *Chinese* is a single category because the majority of participants did not indicate which specific language they had learnt.

As most of the participants had encountered at least one foreign language prior to the experiment. Table A.3 reports which languages they had learnt, and whether they had done so in a classroom or a naturalistic setting.

Appendix B

Instructions Given to Participants

This section reproduces the instructions given to the participants as part of the experiment. They were given in writing and integrated into the computer program. Longer sets of instructions were divided into a number of different screens.

The example of the reproduction task instructions came from Experiment 1, that used the biconditional grammar materials. The other sets of example instructions that have been reproduced here came from the Persian materials.

B.1 Reproduction Task

- In this experiment you will see strings made up of the letters D, F, G, K, L and X.
- After viewing each letter string, you will be asked to type it from memory.
- Be careful never to use capital letters when typing.
- You do not need to type the full stop (period) in the middle of the string.
- Press enter when you have typed all eight letters in the string.
- If you make a mistake, the delete and backspace keys won't work.
- However, you may be able to try again, after viewing the item a second time.
- First you will practise the procedure with some strings made from different letters.
- Although the letters are different, the task is exactly the same.
- Press any button to begin the experiment.

B.2 Source Localisation Task

- In this experiment you will see and hear words and sentences in Persian (but written in the English alphabet — don't worry).
- You will see a picture of the speaker before you hear the sentence.
- If it was the last thing that speaker said, press the green button.
- Otherwise press the red button.
- First you will practise the procedure with some sentences in German.
- Although the language is different, the task is exactly the same.
- Press any button to begin the experiment.

B.3 Timed Test

The timed test was the first element of the procedure for the control participants. Therefore different instructions were used for each group at this point.

B.3.1 Trained Participants

- Now you will be asked some questions.
- You will hear sentences in Persian.
- Some of these are incorrect (they are not actual Persian sentences).
- Please say whether you think they are correct.
- If you are unsure, please guess as well as you can.
- Press green for yes and red for no.
- You will have two seconds after hearing the sentence to answer each question, then the next question will appear.
- After you have heard the sentence the background colour will change to dark blue. This is the time to respond.
- First you will answer two questions in German, to practice the procedure.
- Please take a short break now if you feel you need it.
- Press any button when you are ready to begin this section of the experiment.

B.3.2 Control Participants

First, the control participants were informed verbally that they may have to guess the answers. Then the experiment program began, which started with the following instructions.

- In this experiment you will see and hear sentences in Persian (but written in the English alphabet — don't worry).
- You will be asked questions about these sentences.
- First, you will hear sentences in Persian.
- Some of these are incorrect (they are not actual Persian sentences).
- Please say whether you think they are correct.
- If you are unsure, please guess as well as you can.
- Press green for yes and red for no.
- You will have two seconds after hearing the sentence to answer each question, then the next question will appear.
- After you have heard the sentence the background colour will change to dark blue. This is the time to respond.
- First you will answer two questions in German, to practice the procedure.
- Please take a short break now if you feel you need it.
- Press any button when you are ready to begin this section of the experiment.

B.4 Untimed Test

- Next you will be asked some more questions.
- Again, you will see sentences in Persian. This time you will not hear the sentences.
- Again, press the green button if they are correct and the red button if they are not.
- This time, you can take as long as you like to answer.
- If you are unsure, please guess as well as you can.
- Press any button when you are ready to begin answering (please note, there are no practice sentences in German this time, you will see Persian straight away).

B.5 Correction Test

- You will now be asked twelve more questions.
- You will see a series of incorrect sentences in Persian.
- In each case you will also see four possible reasons why these sentences might be incorrect.
- Press the button with the corresponding number and colour to select which one you think is actually the reason.
- These questions are not timed: take as long as you need to respond.

- If you are unsure, please guess.
- Press any button when you are ready to begin the final stage of the experiment.

APPENDIX C

Calculating the Correction Test Criterion

The trained participants were classified as rule or familiarity detectors or as non-detectors based on their performance in the multiple–choice correction test. If they reached a criterion by responding frequently either on the basis of grammaticality or on the basis of familiarity, they were classified as rule or familiarity detectors respectively. However, if they did not reach the criterion they were classified as non-detectors.

This section describes how the criterion level was calculated to ensure that only approximately 5% of participants could reach it by chance. In each question, the participants had to select the correct answer from a choice of four. One was correct, while the other three were foils. Therefore, there were three ways in which to answer a given question incorrectly. Similarly, there were $3 \times 3 = 9$ ways in which to answer two specific questions incorrectly (three ways to answer the first, and an independent choice of three ways in which to answer the second). In general, the number of ways to answer a specific set of x questions incorrectly was 3^x. This is the first term in the equation.

However, although there were three ways in which to answer a specific question incorrectly, there was more than one question in the test (twelve in the biconditional grammar correction test for example). Naturally therefore, there were twelve items from which to select the one that was answered incorrectly. Thus, the total number of response patterns in which one item was answered incorrectly in the biconditional grammar version was $3 \times 12 = 36$.

The calculation of the second term becomes slightly more complicated when more than one item was answered incorrectly. Specifically, if two items were answered incorrectly there was a choice of twelve in which the first mistake could be made, and eleven in which the second could be made. This gives a total of $12 \times 11 = 132$. However, cases in which Mistake A was made in Item 1 and Mistake B in Item 2, and the patterns where Mistake A was made in Item 2 and Mistake B in Item 1 were actually identical. Therefore the calculation needed to be adjusted. Specifically, the total had to be divided by two, giving $(12 \times 11) \div 2 = 66$. Combining this with the choice of foil within the selected items gives a total of $3^2 \times 66 = 594$ possible response patterns in which two items were answered incorrectly.

If three items were answered incorrectly the first term was equal to 3^3. The selection of items was $12 \times 11 \times 10$, but this time it was divided by $1 \times 2 \times 3$ to give a total of $3^3 \times (12 \times 11 \times 10 \div 1 \times 2 \times 3) = 5940$. In general, the formula in (25) gives the total number of response patterns in which x items are answered incorrectly, where F = number of foils per item and I = number of items.

(25) $F^x \times \frac{I!}{(I-x)! \times x!}$

C.1 Biconditional Grammar Calculation

There were four options in each question and twelve experimental items. Therefore there were $4^{12} = 1677216$ possible response patterns. Table C.1 shows how these were distributed. It indicates that 5.44% included *at least* six correct answers, and therefore that only 5.44% of participants should respond to four or more items correctly by chance. This level was selected as the criterion because it was the closest value to the target of 5%.

C.2 Persian Calculation

There were four possible responses to each question and four experimental items, giving a total number of $4^4 = 256$ response patterns. Following the calculations in Table C.2, thirteen of these (or 5.08%) contained *at least* three correct answers. Therefore only 5.08% of participants could be expected to answer three or more items correctly by chance. This level was selected as the criterion because it was the closest value to the target of 5%.

Errors	Response Patterns		%	Cumulative %
0	3^0	1	0.00	0.00
1	$3^1 \times \frac{12}{1}$	36	0.00	0.00
2	$3^2 \times \frac{12\times11}{1\times2}$	594	0.00	0.00
3	$3^3 \times \frac{12\times11\times10}{1\times2\times3}$	5940	0.04	0.04
4	$3^4 \times \frac{12\times11\times10\times9}{1\times2\times3\times4}$	40095	0.24	0.28
5	$3^5 \times \frac{12\times11\times10\times9\times8}{1\times2\times3\times4\times5}$	192456	1.15	1.43
6	$3^6 \times \frac{12\times11\times10\times9\times8\times7}{1\times2\times3\times4\times5\times6}$	673596	4.01	5.44
7	$3^7 \times \frac{12\times11\times10\times9\times8\times7\times6}{1\times2\times3\times4\times5\times6\times7}$	1732104	10.32	15.76
8	$3^8 \times \frac{12\times11\times10\times9\times8\times7\times6\times5}{1\times2\times3\times4\times5\times6\times7\times8}$	3247695	19.36	35.12
9	$3^9 \times \frac{12\times11\times10\times9\times8\times7\times6\times5\times4}{1\times2\times3\times4\times5\times6\times7\times8\times9}$	4330260	25.81	60.93
10	$3^{10} \times \frac{12\times11\times10\times9\times8\times7\times6\times5\times4\times3}{1\times2\times3\times4\times5\times6\times7\times8\times9\times10}$	3897234	23.23	84.16
11	$3^{11} \times \frac{12\times11\times10\times9\times8\times7\times6\times5\times4\times3\times2}{1\times2\times3\times4\times5\times6\times7\times8\times9\times10\times11}$	2125764	12.67	96.83
12	$3^{12} \times \frac{12\times11\times10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{1\times2\times3\times4\times5\times6\times7\times8\times9\times10\times11\times12}$	531441	3.17	100
TOTAL		**1677216**		

Table C.1: Biconditional grammar correction test criterion. The table illustrates how the criterion was calculated.

Errors	Response Patterns		%	Cumulative %
0	3^0	1	0.39	0.39
1	$3^1 \times \frac{4}{1}$	12	4.69	5.08
2	$3^2 \times \frac{4\times3}{1\times2}$	54	21.09	26.17
3	$3^3 \times \frac{4\times3\times2}{1\times2\times3}$	108	42.19	68.36
4	$3^4 \times \frac{4\times3\times2\times1}{1\times2\times3\times4}$	81	31.64	100
TOTAL		**256**		

Table C.2: Persian correction test criterion. The table illustrates how the criterion was calculated.

Errors	Response Patterns		%	Cumulative %
0	3^0	1	0.02	0.02
1	$3^1 \times \frac{6}{1}$	18	0.44	0.46
2	$3^2 \times \frac{6\times5}{1\times2}$	135	3.30	3.76
3	$3^3 \times \frac{6\times5\times4}{1\times2\times3}$	540	13.18	16.94
4	$3^4 \times \frac{6\times5\times4\times3}{1\times2\times3\times4}$	1215	30.00	46.94
5	$3^5 \times \frac{6\times5\times4\times3\times2}{1\times2\times3\times4\times5}$	1458	35.60	82.54
6	$3^6 \times \frac{6\times5\times4\times3\times2\times1}{1\times2\times3\times4\times5\times6}$	729	17.80	100
TOTAL		**4096**		

Table C.3: Basque correction test criterion. The table illustrates how the criterion was calculated.

C.3 Basque Calculation

There were four options in each question and six experimental items. Therefore there were $4^6 = 4096$ possible response patterns. Table C.3 shows how these were distributed. It indicates that one hundred and fifty–four (or 3.76%) included *at least* four correct answers, and therefore that only 3.76% of participants should respond to four or more items correctly by chance. This level was selected as the criterion because it was the closest value to the target of 5%.

APPENDIX D

Biconditional Grammar Stimuli

Experiment 1 used stimuli generated by a biconditional grammar. Table D.1 shows the twenty–four unique learning items that were extracted from Johnstone and Shanks (2001, Exp. 3). The judgement test items (from the same source) are reproduced in Table D.2.

Learning Items			
DFGD.FDLF	FDLF.DFGD	GXKG.LKXL	LGDL.GLFG
DFGX.FDLK	FDLK.DFGX	GXLG.LKGL	LGXL.GLKG
DFKD.FDXF	FDXK.DFKX	KDFK.XFDX	LKGL.GXLG
DLFD.FGDF	FGLK.DLGX	KDLK.XFGX	LKXL.GXKG
DLGD.FGLF	GDFG.LFDL	KDXK.XFKX	XKDL.KXFG
DXKD.FKXF	GLFG.LGDL	KXLK.XKGX	XLGX.KGLK

Table D.1: Biconditional grammar learning items.

G-F	**G-NF**	**NG-F**	**NG-NF**
DFGL.FDLG	DGKL.FLXG	DFGL.FDLF	DGKL.DLXG
DLKX.FGXK	DXGK.FKLX	DLKX.FDXK	DXGK.FKLD
FDLG.DFGL	FKLX.DXGK	FDLK.DFGL	FKLX.DXGF
FGXF.DLKD	FLXG.DGKL	FGXF.GLKD	FLDG.DGKL
GLKD.LGXK	GDKF.LFXD	GLKD.FGXF	GDKF.LFXG
GXFD.LKDF	GKFX.LXDK	GXKD.LKDF	GKLX.LXDK
KGXK.XLKX	KFLD.XDGF	KDXK.XLKX	DGFL.XLDG
KXFG.XKDL	KGFL.XLDG	GXFG.XKDL	KFXD.XDGF
LKDF.GXFD	LDGK.GFLX	LKDF.GXKD	LDGK.GKLX
LKXF.GXKD	LXDK.GKFX	LKXF.GXKG	LXDK.GKLX
XFDL.KDFG	XDGF.KFLD	XFKX.KGXK	XDGF.KFXD
XFKX.KDXK	XGKF.KLXD	XKDL.KDFG	XDKF.KLXD

Table D.2: Biconditional grammar judgement test items.

Correction Test Items	
DGKL.FLXD	GDKF.LFXG
DKFL.FXDL	GFKL.XDXG
DXGK.FKLD	GKFL.DXDG
FKLX.DXGF	LDGK.GKLX
FXDK.GKFX	XDGF.KFXD
FXLX.DKGF	XDKF.KLXD

Table D.3: Biconditional grammar correction test items.

The correction test items were developed for the current project. All were NG-NF. They are listed in Table D.3.

Appendix E

Persian Stimuli

The same Persian stimuli were used in both Experiments 2 and 3. The learning items are listed in Table E.1.

In the Persian materials, different items were used in each judgement test. Table E.2 shows the stimuli separated between the timed and untimed tests.

Unlike in the biconditional grammar materials, it was possible to generate ungrammatical fillers for the Persian correction test. The fillers either had errors in the word order or the verbal agreement morphology was missing. The experimental items had incorrect number agreement on the verb. The full list can be seen in Table E.3

Persian	Translation
Mehmunane shohar keshad	The guests of the cousin left
Mehmunane shoharan keshad	The guests of the cousins left
Mehmunane shoharan tarsad	The guests of the cousins arrived
Mehmunane yateem keshad	The guests of the doctor left
Mehmunane yateeman geerad	The guests of the doctors ran
Mehmunane yateeman keshad	The guests of the doctors left
Mehmunane yateeman tarsad	The guests of the doctors arrived
Mehmune shohar geere	The guest of the cousin ran
Mehmune shohar keshe	The guest of the cousin left
Mehmune shoharan keshe	The guest of the cousins left
Mehmune yateem geere	The guest of the doctor ran
Mehmune yateem keshe	The guest of the doctor left
Mehmune yateem tarse	The guest of the doctor arrived
Mehmune yateeman keshe	The guest of the doctors left
Shoharane mehmun keshad	The cousins of the guest left
Shoharane mehmunan geerad	The cousins of the guests ran
Shoharane mehmunan keshad	The cousins of the guests left
Shoharane yateem keshad	The cousins of the doctor left
Shoharane yateeman geerad	The cousins of the doctors ran
Shoharane yateeman keshad	The cousins of the doctors left
Shoharane yateeman tarsad	The cousins of the doctors arrived
Shohare mehmun keshe	The cousin of the guest left
Shohare mehmun tarse	The cousin of the guest arrived
Shohare mehmunan keshe	The cousin of the guests left
Shohare yateem geere	The cousin of the doctor ran
Shohare yateem keshe	The cousin of the doctor left
Shohare yateem tarse	The cousin of the doctor arrived
Shohare yateeman keshe	The cousin of the doctors left
Yateemane mehmun keshad	The doctors of the guest left
Yateemane mehmunan geerad	The doctors of the guests ran
Yateemane mehmunan keshad	The doctors of the guests arrived
Yateemane shohar keshad	The doctors of the cousin left
Yateemane shoharan keshad	The doctors of the cousins ran
Yateemane shoharan tarsad	The doctors of the cousins arrived
Yateeme mehmun keshe	The doctor of the guest left
Yateeme mehmun tarse	The doctor of the guest arrived
Yateeme mehmunan keshe	The doctor of the guests left
Yateeme shohar geere	The doctor of the cousin ran
Yateeme shohar keshe	The doctor of the cousin left
Yateeme shoharan keshe	The doctor of the husbands left

Table E.1: Persian learning items.

Timed Test		Untimed Test	
Item	**Type**	**Item**	**Type**
Shohare yateem geere	G-F	Yateeme mehmun tarse	G-F
Shohare mehmun tarse	G-F	Mehmune yateem tarse	G-F
Mehmune shohar geere	G-F	Shoharane mehmunan geerad	G-F
Yateemane mehmunan geerad	G-F	Mehmunane yateeman geerad	G-F
Shoharane yateeman geerad	G-F	Mehmunane yateeman tarsad	G-F
Shoharane yateeman tarsad	G-F	Mehmunane shoharan tarsad	G-F
Yateeme mehmun geere	G-NF	Yateeme shoharan geere	G-NF
Shohare mehmun geere	G-NF	Yateeme mehmunan tarse	G-NF
Yateeme shoharan tarse	G-NF	Shohare yateeman geere	G-NF
Yateeme mehmunan geere	G-NF	Shohare mehmunan tarse	G-NF
Shohare yateeman tarse	G-NF	Mehmune yateeman geere	G-NF
Shohare mehmunan geere	G-NF	Mehmune shoharan geere	G-NF
Mehmune yateeman tarse	G-NF	Yateemane shohar tarsad	G-NF
Mehmune shoharan tarse	G-NF	Yateemane mehmun geerad	G-NF
Yateemane shohar geerad	G-NF	Shoharane yateem geerad	G-NF
Yateemane mehmun tarsad	G-NF	Shoharane mehmun geerad	G-NF
Shoharane yateem tarsad	G-NF	Mehmunane yateem geerad	G-NF
Shoharane mehmun tarsad	G-NF	Mehmunane shohar tarsad	G-NF
Mehmunane yateem tarsad	G-NF	Yateemane mehmunan tarsad	G-NF
Mehmunane shohar geerad	G-NF	Shoharane mehmunan tarsad	G-NF
Shoharane yateem geere	NG-F	Yateemane shohar geere	NG-F
Shoharane yateem tarse	NG-F	Shoharane mehmun tarse	NG-F
Mehmunane shohar geere	NG-F	Mehmunane yateem geere	NG-F
Yateeme shoharan tarsad	NG-F	Mehmunane yateem tarse	NG-F
Shohare mehmunan geerad	NG-F	Shohare yateeman geerad	NG-F
Mehmune yateeman tarsad	NG-F	Mehmune shoharan tarsad	NG-F
Yateeme shohar geerad	NG-NF	Yateeme shohar tarsad	NG-NF
Yateeme mehmun tarsad	NG-NF	Yateeme mehmun geerad	NG-NF
Shohare yateem tarsad	NG-NF	Shohare yateem geerad	NG-NF
Shohare mehmun tarsad	NG-NF	Shohare mehmun geerad	NG-NF
Mehmune yateem tarsad	NG-NF	Mehmune yateem geerad	NG-NF
Mehmune shohar geerad	NG-NF	Mehmune shohar tarsad	NG-NF
Yateemane mehmun geere	NG-NF	Yateeme mehmunan tarsad	NG-NF
Shoharane mehmun geere	NG-NF	Shohare mehmunan tarsad	NG-NF
Yateemane shoharan tarse	NG-NF	Yateemane shoharan geere	NG-NF
Yateemane mehmunan geere	NG-NF	Yateemane mehmunan tarse	NG-NF
Shoharane yateeman tarse	NG-NF	Shoharane yateeman geere	NG-NF
Shoharane mehmunan geere	NG-NF	Shoharane mehmunan tarse	NG-NF
Mehmunane yateeman tarse	NG-NF	Mehmunane yateeman geere	NG-NF
Mehmunane shoharan tarse	NG-NF	Mehmunane shoharan geere	NG-NF

Table E.2: Persian judgement test items.

Item	Type	Item	Type
Mehmunane shohar keshe	Exp	Mehmune geere shoharan	Filler
Mehmune shoharan tarsad	Exp	Shoharane mehmun geer	Filler
Mehmune yateeman tarsad	Exp	Shoharane mehmunane geerad	Filler
Shoharane yateeman tarse	Exp	Tarsad shoharane yateem	Filler
Geere mehmun shoharan	Filler	Shohare mehmun tars	Filler
Mehmun shoharan geere	Filler	Yateem geere mehmunan	Filler

Table E.3: Persian correction test items.

Appendix F

Basque Stimuli

F.1 Experiment 4

Basque	Translation
Motilye gisona ekarri du	The boy found the man
Motilye lahuna ahurtu du	The boy followed the girl
Motilye gisonak ikusi deetu	The boy saw the men
Motilye lahunak ekarri deetu	The boy found the girls
Gisone motilya ahurtu du	The man followed the boy
Gisone lahuna ekarri du	The man found the girl
Gisone motilyak ikusi deetu	The man saw the boys
Gisone lahunak ahurtu deetu	The man followed the girls
Lahune gisona ekarri du	The girl found the man
Lahune motilya ikusi du	The girl saw the boy
Lahune gisonak ikusi deetu	The girl saw the men
Lahune motilyak ahurtu deetu	The girl followed the boys
Motilye gisona ikusi du	The boy saw the man
Motilye lahuna ekarri du	The boy found the girl
Motilye gisonak ekarri deetu	The boy found the men
Motilye lahunak ahurtu deetu	The boy followed the girls
Gisone motilya ikusi du	The man saw the boy
Gisone lahuna ahurtu du	The man followed the girl
Gisone motilyak ahurtu deetu	The man followed the boys
Gisone lahunak ekarri deetu	The man found the girls
Lahune gisona ikusi du	The girl saw the man
Lahune motilya ahurtu du	The girl followed the boy
Lahune gisonak ekarri deetu	The girl found the men
Lahune motilyak ikusi deetu	The girl saw the boys

Table F.1: Basque learning items from Experiment 4.

In the stimuli employed in Experiment 4, the number value of the object was varied. Singular subjects were employed in the versions reproduced here.

However for half of the participants all of the subject nouns were plural and therefore suffixed with *-k*, while there was a *-te* suffix on the auxiliary (the final word). There was no other difference between the two sets of stimuli.

Item	Type	Item	Type
Motilye gisona ekarri du	G-F	Motilye gisona ekarri deetu	NG-F
Motilye lahunak ekarri deetu	G-F	Motilye lahunak ekarri du	NG-F
Gisone motilya ahurtu du	G-F	Gisone motilya ahurtu deetu	NG-F
Gisone lahunak ahurtu deetu	G-F	Gisone lahunak ahurtu du	NG-F
Lahune motilya ikusi du	G-F	Lahune motilya ikusi deetu	NG-F
Lahune gisonak ikusi deetu	G-F	Lahune gisonak ikusi du	NG-F
Motilye lahuna ekarri du	G-F	Motilye lahuna ekarri deetu	NG-F
Motilye gisonak ekarri deetu	G-F	Motilye gisonak ekarri du	NG-F
Gisone lahuna ahurtu du	G-F	Gisone lahuna ahurtu deetu	NG-F
Gisone motilyak ahurtu deetu	G-F	Gisone motilyak ahurtu du	NG-F
Lahune gisona ikusi du	G-F	Lahune gisona ikusi deetu	NG-F
Lahune motilyak ikusi deetu	G-F	Lahune motilyak ikusi du	NG-F
Motilye gisona ahurtu du	G-NF	Motilye gisona ahurtu deetu	NG-NF
Motilye lahuna ikusi du	G-NF	Motilye lahuna ikusi deetu	NG-NF
Motilye gisonak ahurtu deetu	G-NF	Motilye gisonak ahurtu du	NG-NF
Motilye lahunak ikusi deetu	G-NF	Motilye lahunak ikusi du	NG-NF
Gisone motilya ekarri du	G-NF	Gisone motilya ekarri deetu	NG-NF
Gisone lahuna ikusi du	G-NF	Gisone lahuna ikusi deetu	NG-NF
Gisone motilyak ekarri deetu	G-NF	Gisone motilyak ekarri du	NG-NF
Gisone lahunak ikusi deetu	G-NF	Gisone lahunak ikusi du	NG-NF
Lahune gisona ahurtu du	G-NF	Lahune gisona ahurtu deetu	NG-NF
Lahune motilya ekarri du	G-NF	Lahune motilya ekarri deetu	NG-NF
Lahune gisonak ahurtu deetu	G-NF	Lahune gisonak ahurtu du	NG-NF
Lahune motilyak ekarri deetu	G-NF	Lahune motilyak ekarri du	NG-NF

Table F.2: Basque judgement test items from Experiment 4.

Table F.2 shows the judgement test items from Experiment 4. The same stimuli were used for both the timed and the untimed tests.

The six experimental and six filler items used in the sentence correction test are listed in Table F.3. All of the items were NG-NS. The fillers were ungrammatical as a result of word order errors.

F.2 Experiment 5

In Experiment 5 the number value of the subject was varied, while that of the object was held constant. The versions included here use singular objects. The plural versions were identical, except that plurality was marked by a *-k* suffix on

Item	Type
Motilye gisona ahurtu deetu	Exp
Motilye lahunak ikusi du	Exp
Gisone lahuna ikusi deetu	Exp
Gisone motilyak ekarri du	Exp
Lahune motilya ekarri deetu	Exp
Lahune gisonak ahurtu du	Exp
Motilye lahunak deetu ekarri	Filler
Lahune deetu gisonak ikusi	Filler
Deetu gisone motilyak ahurtu	Filler
Gisone ekarri lahuna du	Filler
Ikusi motilye gisona du	Filler
Du ahurtu motilya lahune	Filler

Table F.3: Basque correction test items from Experiment 4.

every object noun and the first morpheme in the auxiliary was *deetu* rather than *du*. Table F.4 shows the learning items, Table F.5 holds the judgement test items, while the correction test items can be seen in Table F.6.

Basque	Translation
Gisone lahuna ahurtu du	The man followed the girl
Gisone lahuna ekarri du	The man found the girl
Gisone motilya ahurtu du	The man followed the boy
Gisone motilya ikusi du	The man saw the boy
Gisonek lahuna ahurtu dute	The men followed the girl
Gisonek lahuna ekarri dute	The men found the girl
Gisonek motilya ahurtu dute	The men followed the boy
Gisonek motilya ikusi dute	The men saw the boy
Lahune gisona ekarri du	The girl found the man
Lahune gisona ikusi du	The girl saw the man
Lahune motilya ahurtu du	The girl followed the boy
Lahune motilya ikusi du	The girl saw the boy
Lahunek gisona ekarri dute	The girls found the man
Lahunek gisona ikusi dute	The girls saw the man
Lahunek motilya ahurtu dute	The girls followed the boy
Lahunek motilya ikusi dute	The girls saw the boy
Motilye gisona ekarri du	The boy found the man
Motilye gisona ikusi du	The boy saw the man
Motilye lahuna ahurtu du	The boy followed the girl
Motilye lahuna ekarri du	The boy found the girl
Motilyek gisona ekarri dute	The boys found the man
Motilyek gisona ikusi dute	The boys saw the man
Motilyek lahuna ahurtu dute	The boys followed the girl
Motilyek lahuna ekarri dute	The boys found the girl

Table F.4: Basque learning items from Experiment 5.

Item	Type	Item	Type
Motilye gisona ekarri du	G-F	Motilye gisona ekarri dute	NG-F
Motilyek lahuna ekarri dute	G-F	Motilyek lahuna ekarri du	NG-F
Gisone motilya ahurtu du	G-F	Gisone motilya ahurtu dute	NG-F
Gisonek lahuna ahurtu dute	G-F	Gisonek lahuna ahurtu du	NG-F
Lahune motilya ikusi du	G-F	Lahune motilya ikusi dute	NG-F
Lahunek gisona ikusi dute	G-F	Lahunek gisona ikusi du	NG-F
Motilye lahuna ekarri du	G-F	Motilye lahuna ekarri dute	NG-F
Motilyek gisona ekarri dute	G-F	Motilyek gisona ekarri du	NG-F
Gisone lahuna ahurtu du	G-F	Gisone lahuna ahurtu dute	NG-F
Gisonek motilya ahurtu dute	G-F	Gisonek motilya ahurtu du	NG-F
Lahune gisona ikusi du	G-F	Lahune gisona ikusi dute	NG-F
Lahunek motilya ikusi dute	G-F	Lahunek motilya ikusi du	NG-F
Motilye gisona ahurtu du	G-NF	Motilye gisona ahurtu dute	NG-NF
Motilye lahuna ikusi du	G-NF	Motilye lahuna ikusi dute	NG-NF
Motilyek gisona ahurtu dute	G-NF	Motilyek gisona ahurtu du	NG-NF
Motilyek lahuna ikusi dute	G-NF	Motilyek lahuna ikusi du	NG-NF
Gisone motilya ekarri du	G-NF	Gisone motilya ekarri dute	NG-NF
Gisone lahuna ikusi du	G-NF	Gisone lahuna ikusi dute	NG-NF
Gisonek motilya ekarri dute	G-NF	Gisonek motilya ekarri du	NG-NF
Gisonek lahuna ikusi dute	G-NF	Gisonek lahuna ikusi du	NG-NF
Lahune gisona ahurtu du	G-NF	Lahune gisona ahurtu dute	NG-NF
Lahune motilya ekarri du	G-NF	Lahune motilya ekarri dute	NG-NF
Lahunek gisona ahurtu dute	G-NF	Lahunek gisona ahurtu du	NG-NF
Lahunek motilya ekarri dute	G-NF	Lahunek motilya ekarri du	NG-NF

Table F.5: Basque judgement test items from Experiment 5.

Item	Type
Motilye gisona ahurtu dute	Exp
Motilyek lahuna ikusi du	Exp
Gisone lahuna ikusi dute	Exp
Gisonek motilya ekarri du	Exp
Lahune motilya ekarri dute	Exp
Lahunek gisona ahurtu du	Exp
Motilyek lahuna dute ekarri	Filler
Lahunek dute gisona ikusi	Filler
Dute gisonek motilya ahurtu	Filler
Gisone ekarri lahuna du	Filler
Ikusi motilye gisona du	Filler
Du ahurtu motilya lahune	Filler

Table F.6: Basque correction test items from Experiment 5.

APPENDIX G

Simulation Input

G.1 Simulation Input in the Reproduction Task

Simulations 1 and 2 modelled experiments that used the reproduction task. In this procedure, the learners were exposed to a single item before retyping it from memory. The network's input included the learners' productions interspersed with the learning items. Table G.1 shows a sample of the input provided in Simulation 1a, in which the network was trained on the biconditional stimuli using the equivalent of the reproduction task. An active node had a value of one and an inactive one of zero. Note that the network's training task was next–letter prediction with both the learning and reproduced items.

G.2 Simulation Input in the Source Localisation Task

The experiments using the source localisation task were also modelled. In this task, the learners were exposed to four item–speaker pairs, before verifying whether a fifth pair had been encountered in the immediately preceding set of four items. Therefore the network's input included the items from these source localisation questions interspersed with the learning items. Table G.2 shows a sample of the input provided in Simulation 3a, in which the network was trained on the Persian stimuli using the equivalent of the source localisation task. Note that the network's training task was next–word prediction with both the learning and source localisation items.

Units													Item	
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**Letter**	**Type**
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	Input
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	Reproduction
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	1	1	0	0	0	0	0	0	0	L	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	Input
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	Reproduction
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	1	1	0	X	
0	0	0	0	0	0	1	1	0	0	0	0	0	K	
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	
0	0	1	1	0	0	0	0	0	0	0	0	0	F	
1	1	0	0	0	0	0	0	0	0	0	0	0	D	
0	0	0	0	0	0	0	0	1	1	0	0	0	L	
0	0	1	1	0	0	0	0	0	0	0	0	0	F	Input
1	1	0	0	0	0	0	0	0	0	0	0	0	D	
0	0	1	1	0	0	0	0	0	0	0	0	0	F	
0	0	0	0	1	1	0	0	0	0	0	0	0	G	
1	1	0	0	0	0	0	0	0	0	0	0	0	D	

Table G.1: Sample input for the learning phase of Simulation 1a. The learning items are labelled *input*, while the *reproduction* items are the learners' attempts to retype the previous stimulus in the reproduction task.

Units													Item	
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**Letter**	**Type**
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmune	
0	0	1	0	0	0	0	0	0	1	0	0	0	Yateeman	
0	0	0	0	1	0	0	0	0	0	1	0	0	Keshe	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
0	0	1	0	0	0	0	0	0	1	0	0	0	Yateemane	
1	0	0	0	0	0	0	0	0	1	0	0	0	Mehmunan	
0	0	0	0	1	0	0	0	0	0	0	1	0	Keshad	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmune	
0	0	1	0	0	0	0	0	1	0	0	0	0	Yateem	
0	0	0	0	1	0	0	0	0	0	1	0	0	Keshe	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
0	0	1	0	0	0	0	0	1	0	0	0	0	Yateeme	
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmun	
0	0	0	0	0	1	0	0	0	0	1	0	0	Tarse	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Source
0	1	0	0	0	0	0	0	1	0	0	0	0	Shohare	
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmun	
0	0	0	0	1	0	0	0	0	0	1	0	0	Keshe	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmune	
0	0	1	0	0	0	0	0	1	0	0	0	0	Yateem	
0	0	0	1	0	0	0	0	0	0	1	0	0	Geere	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
0	1	0	0	0	0	0	0	1	0	0	0	0	Shohare	
0	0	1	0	0	0	0	0	1	0	0	0	0	Yateem	
0	0	0	0	0	1	0	0	0	0	1	0	0	Tarse	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
0	1	0	0	0	0	0	0	0	1	0	0	0	Shoharane	
0	0	1	0	0	0	0	0	0	1	0	0	0	Yateeman	
0	0	0	1	0	0	0	0	0	0	0	1	0	Geerad	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Input
0	1	0	0	0	0	0	0	1	0	0	0	0	Shohare	
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmun	
0	0	0	0	1	0	0	0	0	0	1	0	0	Keshe	
0	0	0	0	0	0	0	0	0	0	0	0	1	Stop	Source
1	0	0	0	0	0	0	0	1	0	0	0	0	Mehmune	
0	1	0	0	0	0	0	0	1	0	0	0	0	Shohar	
0	0	0	0	1	0	0	0	0	0	1	0	0	Keshe	

Table G.2: Sample input for the learning phase of Simulation 3a. The learning items are labelled *input*, while the *source* items were presented for the questions in the source localisation task.

References

Alavi, B., & Lorenz, M. (1999). *Persisch: Ein Standardwerk für Anfänger* [Persian: A standard work for beginners]. Berlin: Langenscheidts.

Altmann, G. T. M., & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science*, *284*, 875a-876a.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 899-912.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321-324.

Atkinson, A. P., Thomas, M. S. C., & Cleeremans, A. (2000). Consciousness: Mapping the theoretical landscape. *Trends in Cognitive Sciences*, *4*, 372-382.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340-357.

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory*, *25*, 657689.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum Associates.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, *29*, 81-103.

Bitan, T., & Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research*, *19*, 229-243.

Bock, J. K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*, 437-458.

Bock, J. K., & Griffen, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General*, *129*, 177-192.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45-93.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*, 451-459.

Boyland, J. T., & Anderson, J. R. (1998). Evidence that syntactic priming is long-lasting. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316-323.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, *24*, 215-221.

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

Channon, S., Shanks, D. R., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, *40*, 2185-2197.

Christiansen, M. H., & Ellefson, M. R. (2002). Linguistic adaptation without linguistic constraints. In A. Wray (Ed.), *The transition to language* (p. 335-358). Oxford: Oxford University Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406-416.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making* (p. 1-40). Hove: Psychology Press Ltd.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235-353.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 17-30.

Coppieters, R. (1987). Competence differences between native and fluent nonnative speakers. *Language*, *63*, 544-573.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, *9*, 522-533.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 189-202.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workshop 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*(3), 379-410.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphoyntax. *Studies in Second Language Acquisition*, *19*, 195-221.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language acquisition* (p. 125-151). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (p. 313-348). Oxford: Blackwell.

Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, *79*, 221-237.

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science*, *16*, 41-79.

Dienes, Z., Altmann, G., Gao, S.-J., & Goode, A. (1995). The transfer of implicit knowledge across domains. *Language and Cognitive Processes*, *10*, 363-367.

Dienes, Z., Altmann, G. T. M., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science*, *23*, 53-82.

Dienes, Z., Broadbent, D., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 875-887.

Dienes, Z., & Perner, J. (1994). Dissociable definitions of consciousness. *Behavioral and Brain Sciences*, *17*, 403-404.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735808.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*, 338-351.

Don, A. J., Schellenberg, E. G., Reber, A. S., DiGirolamo, K. M., & Wang, P. P. (2003). Implicit learning in children and adults with Williams Syndrome. *Developmental Neuropsychology*, *23*, 201-225.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, *113*, 541-555.

Eberhard, K. M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, *36*, 147-164.

Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, *5*, 289-318.

Ellis, N. C. (2002a). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143-188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 297-339.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (p. 63-103). Oxford: Blackwell.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305-352.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1-24.

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*, 145-171.

Ellis, N. C., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? the frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*, 307-336.

Ellis, R. (1994). Second language acquisition research: An overview. In *The study of second language acquisition* (p. 11-39). Oxford: Oxford University Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141-172.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*, 431-463.

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005). *Studies in Second Language Acquisition*, *29*, 119-126.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Erlam, R. (2003). The effects of deductive and inductive instruction on the acquisition of direct object pronouns in French as a second language. *Modern Language Journal*, *87*, 242-260.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464-491.

Erlam, R., Ellis, R., & Loewen, S. (2007, September). *Is Incidental Learning Possible in the Classroom? The Case of Third Person -s*. Paper presented at the 40th Annual Meeting of the British Association for Applied Linguistics, University of Edinburgh, UK.

Ferreira, V. S., Bock, J. K., Wilson, M. P., & Cohen, N. J. (2008). Memory for syntax despite amnesia. *Psychological Science*, *19*, 940-946.

Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes*, *21*, 1011-1029.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.

Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 567-584.

Frensch, P. A., & Miner, C. S. (1994). Effects of presention rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition*, *22*, 95-110.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 2458-2463.

Frith, C., Perry, R., & Lumer, E. (1999). The neural correlates of conscious experience: An experimental framework. *Trends in Cognitive Sciences*, *3*, 105-114.

Gaillard, V., Vandenberghe, M., Destrebecqz, A., & Cleeremans, A. (2006). First- and third-person approaches in implicit learning research. *Consciousness and Cognition*, *15*, 709-722.

Gass, S. M., & Mackey, A. (2002). Frequency effects and second language acquisition. *Studies in Second Language Acquisition*, *24*, 249-260.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of l2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1-50.

Gómez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*, 154-207.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431-436.

Gómez, R. L. (2008, September). *Constraining Children's Generalizations: Processing Constraints on Language Learning*. Paper presented at the 14th Annual Conference on Architectures and Mechanisms for Language Processing, Cambridge, UK.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109-135.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178-186.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*, 567580.

Gómez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 396-410.

Green, R. E. A., & Shanks, D. R. (1993). On the existence of independent explicit and implicit learning systems: An examination of some evidence. *Memory and Cognition*, *21*, 304-317.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*, 1-23.

Hartsuiker, R. J., Antón-Méndez, I., & Zee, M. van. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language*, *45*, 546-572.

Hauser, E. (2000). Explicit and incidental instruction and learner awareness. In B. Swierzbin, F. Morris, M. Anderson, C. Klee, & E. Tarone (Eds.), *Social and cognitive factors in second language acquisition* (p. 326-344). Somerville, MA: Cascadilla Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1579.

Hebb, D. O. (2002). *The organization of behavior* (2nd ed.). Mahwah, NJ.: Lawrence Erlbaum Associates.

Howard, J. H., Mutter, S. A., & Howard, D. V. (1992). Serial pattern learning by event observation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 1029-1039.

Hulstijn, J. H. (2002a). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193223.

Hulstijn, J. H. (2002b). What does the impact of frequency tell us about the language acquisition device? *Studies in Second Language Acquisition*, *24*, 269-273.

Hulstijn, J. H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, *31*, 23-43.

Isemonger, I. M. (2007). Operational definitions of explicit and implicit knowledge. *Studies in Second Language Acquisition*, *29*, 101-118.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

James, G., Burley, D., Clements, D., Dyke, P., Searle, J., & Wright, J. (1996). *Modern engineering mathematics*. Harlow, UK: Addison-Wesley.

Jansen, W. (2002). *Beginner's Basque*. New York: Hippocrene Books.

Jarvis, S., & Odlin, T. (2000). Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition*, *22*, 535-556.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology A*, *54*, 1105-1124.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 236-259.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 948-969.

Jiménez, L., & Vazquez, G. A. (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological Research*, *69*, 352-368.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology*, *42*, 61-112.

Kanno, K. (1997). The acquisition of null and overt pronominals in Japanese by English speakers. *Second Language Research*, *13*, 265-287.

Kemp, C. (2001). *Metalinguistic awareness in multilinguals: Implicit and explicit grammatical awareness and its relationship with language experience and language attainment*. Unpublished doctoral dissertation, University of Edinburgh.

Kersten, A. W., & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, *44*, 250-273.

Klapper, J., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research*, *7*(3), 285-314.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Lardiere, D. (1998). Case and tense in the "fossilized" steady state. *Second Language Research*, *14*, 1-26.

Lieberman, M., Chang, G., Chiao, J., Bookheimer, S., & Knowlton, B. (2004). An event-related fMRI study of artificial grammar: Learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, *16*, 427-438.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (p. 487-535). Oxford: Blackwell.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maia, T. V., & Cleeremans, A. (2005). Consciousness: Converging insights from connectionist modeling and neuroscience. *Trends in Cognitive Sciences*, *9*, 397-404.

Manza, L., & Bornstein, R. F. (1995). Affective discrimination and the implicit learning process. *Consciousness and Cognition*, *4*, 399-409.

Marcus, G. F. (1999a). Reply to Seidenberg and Elman. *Trends in Cognitive Sciences*, *3*, 289.

Marcus, G. F. (1999b). Response: Rule learning by seven-month-old infant and neural networks. *Science*, *284*, 875a-876a.

Mathews, R., Buss, R., Stanley, W., Blanchard-Fields, F., Cho, J., & Druhan, B. (1989). 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. L. (1998). Complementary learning systems in the brain: A connectionist approach to explicit and implicit cognition and memory. *Annals of the New York Academy of Sciences*, *843*, 153-169.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419-457.

Meulemans, T., & Linden, M. Van der. (2002). Artificial grammar learning in amnesia. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making.* (p. 144-163). Hove: Psychology Press Ltd.

Michas, I. C., & Berry, D. C. (1994). Implicit and explicit processes in a second-language learning task. *European Journal of Cognitive Psychology*, *6*, 357-381.

Morgan-Short, K. (2007). *A neurolinguistic investigation of late-learned second language: Effects of explicit and implicit conditions*. Unpublished doctoral dissertation, Georgetown University.

Moshiri, L. (1988). *Colloquial Persian*. London: Routledge.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of american sign language. *Language Sciences*, *10*, 147-172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11-28.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*, 85-117.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*, 569-587.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.

Nissen, M. J., Willingham, D., & Hartman, M. (1989). Explicit and implicit remembering: When is learning preserved in amnesia? *Neuropsychologia*, *27*, 341-352.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and qualitative meta-analysis. *Language Learning*, *50*, 417-528.

O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*, *22*, 127-196.

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (p. 436-486). Oxford: Blackwell.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427-456.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604-607.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General*, *119*, 264-275.

Poletiek, F. H. (2002). Implicit learning of a recursive rule in an artificial grammar. *Acta Psychologica*, *111*, 323-335.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855-863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115-119.

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.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219-235.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, *6*, 189-221.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 492-502.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, *5*, 333-361.

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, *10*, 248-263.

Reed, J., & Johnson, P. (1994a). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 585-594.

Reed, J., & Johnson, P. (1994b). New evidence for unconscious sequence learning. *Behavioral and Brain Sciences*, *17*, 419-420.

Ringbom, H. (1992). On L1 transfer in L2 comprehension and L2 production. *Language Learning*, *42*, 85-112.

Robinson, P. J. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27-67.

Robinson, P. J. (1997a). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*, 223-247.

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

Robinson, P. J. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning. *Studies in Second Language Acquisition*, *27*, 235-268.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67-109.

Rowland, L. A., & Shanks, D. R. (2006). Sequence learning and selection difficulty. *Journal of Experimental Psychology: Human Learning and Memory*, *32*, 287-299.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493-515.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110-113.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27-52.

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.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101-105.

Samareh, Y. (1993a). *Persian language teaching (AZFA): Book 1*. Tehran: Alhoda.

Samareh, Y. (1993b). *Persian language teaching (AZFA): Book 2*. Tehran: Alhoda.

Santesteban, M. (2008, July). *Encoding Number Agreement in L2: Evidence from Subject-Verb and Object-Verb Agreement in Basque (L2)*. Paper presented at the Development Linguistics Research Group, University of Edinburgh, UK.

Schacter, D. L., & Graf, P. (1989). Modality specificity of implicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 3-12.

Scheffler, P., & Cinciala, M. (2008, September). *Explicit L2 Grammar Knowledge and Language Use*. Paper presented at the 18th European Second Language Association Conference, University of Provence, France.

Schmidt, R. W. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206-226.

Schmidtke, V., & Heuer, H. (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research*, *60*, 53-71.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (p. 1-22). Amsterdam: John Benjamins.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime users guide*. Pittsburgh, PA: Psychology Software Tools.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research*, *61*, 175-190.

Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer / full access model. *Second Language Research*, *12*, 40-72.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569-588.

Seliger, H. W. (1979). On the nature and function of language rules in language teaching. *TESOL Quarterly*, *13*, 359-369.

Shanks, D. R., & Johnstone, T. (2002). Modularity and artificial grammar learning. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making.* (p. 93-119). Hove: Psychology Press Ltd.

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

Shanks, D. R., Rowland, L. A., & Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research*, *69*, 369-382.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367-447.

Sinnett, S., Costa, A., & Soto-Faraco, S. (2006). Manipulating inattentional blindness within and across sensory modalities. *Quarterly Journal of Experimental Psychology*, *59*, 1425-1442.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 1521-1543.

Squire, L. R., Hamann, S., & Knowlton, B. (1994). Dissociable learning and memory systems of the brain. *Behavioral and Brain Sciences*, *17*, 422-423.

St. John, M. F., & Shanks, D. R. (1997). Implicit learning from an information processing standpoint. In D. C. Berry (Ed.), *How implicit is implicit learning* (p. 162-194). Oxford: Oxford University Press.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706-716.

Tkachenko, E. (2007, September). *The Role of Type and Token Input Frequencies in the L2 Acquisition of Verbal Morphology*. Paper presented at the 17th European Second Language Association Conference, Newcastle, UK.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183-203.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25-B34.

Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, *26*, 479-504.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (p. 153-172). Amsterdam: John Benjamins.

Vries, M. H. de, Barth, A., Knecht, S., Zwitserlood, P., & Floeel, A. (2008, September). *Stimulation of Broca's Area by Anodal tDCS Enhances Artificial Grammar Learning: A Pilot Study*. Poster presented at the 14th Annual Conference on Architectures and Mechanisms for Language Processing, Cambridge, UK.

White, L. (2003). On the nature of interlanguage representation: Universal grammar in the second language. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (p. 19-42). Oxford: Blackwell.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227-248.

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

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269-304.

Williams, J. N., & Lovatt, P. (2005). Phonological memory and rule learning. *Language Learning*, *55*, 177-233.

Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science*, *10*, 531-534.

Wray, A. (2004). "Here's one i prepared earlier": Formulaic language on television. In N. Schmitt (Ed.), *Formulaic sequences* (p. 249-268). Amsterdam: John Benjamins.