DEVELOPING IMPLICIT AND EXPLICIT KNOWLEDGE OF L2 CASE MARKING UNDER INCIDENTAL LEARNING CONDITIONS

Robert Johnson Rogers

UCL Institute of Education

Thesis submitted for the degree of Doctor of Philosophy

2016

‘I, Robert Johnson Rogers, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.’

Word count (exclusive of front matter, reference list and appendices): 79,987

Signature --

ABSTRACT

This thesis reports on five experiments that investigated the role of awareness in second language acquisition (SLA). Previous research has demonstrated that various areas of second language (L2) grammar can be acquired as result of incidental exposure. In addition, it has been shown that this exposure can lead to implicit knowledge. The results from recent experiments examining the incidental learning of L2 morphology, however, have found little evidence of incidental learning. Furthermore, what little research there is has not provided firm evidence of the nature of the knowledge acquired in these experiments. This thesis set out to address these gaps by testing the degree to which L2 inflectional morphology can be acquired as a result of incidental exposure, and whether the resulting knowledge is implicit or explicit in nature.

The experiments in this thesis followed two different methodologies to address the issues outlined above. The first four experiments followed the artificial grammar learning paradigm, and the final experiment followed a self-paced reading methodology. In the training phase of all experiments, participants were exposed to an artificial language system based on Czech morphology under incidental learning conditions. Subjective measures of awareness and retrospective verbal reports were used to address awareness in all five experiments. The results of the first four experiments indicated that L2 case markers can be learned in the presence of low levels of awareness (below the threshold of verbalisation). These results were supported by the findings of the final experiment. However, the final experiment also revealed that this knowledge can be utilised automatically. Thus, the results of this thesis provide additional data to address the aspects of L2 grammar that can be acquired as a result of incidental exposure and provide further insight into the characteristics of the knowledge that is acquired as a result of this exposure.

DISSEMINATION

Peer-reviewed publications

1. Data from Pilot Study 1, Pilot Study 2, and Experiment 1 have been published as a book chapter: Rogers, J., Révész, A., & Rebuschat, P. (2015). Challenges in implicit learning research: Validating a novel artificial language. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 273-298), Amsterdam: John Benjamins.

2. Data from Experiment 1 have been accepted for publication as a journal article: Rogers, J., Révész, A., & Rebuschat, P. (in press). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics.*

3. Data from Experiment 2 is currently under review for publication as a journal article: Rogers, J. (under review). Explicit knowledge of second language case markings.

4. A manuscript including data from Experiment 3 is currently in preparation.

Conference presentations and seminars

1. Data from Pilot Study 1, Pilot Study 2, and Experiment 1 were presented at EUROSLA 23. Title: Implicit Learning of L2 Grammar: An investigation into Czech Morphology. Amsterdam: August 30th, 2013.

2. Data from Pilot Study 1, Pilot Study 2, and Experiment 1 were presented to the SLLAT Research Group at Lancaster University. Title: Implicit Learning of L2 Morphology. Lancaster: October 16th, 2013.

3. Data from Experiment 2 were presented at AAAL 2014. Title: Incidental Exposure and the Implicit Learning of L2 Case-Markings. Portland: March 22nd, 2014.

4. Data from Experiment 3 were presented as part of the PhD Seminar Series, IOE. Title: Incidental Exposure and the Development of Unconscious Knowledge of L2 morphology. London: January 27th, 2014.

5. Data from Experiment 3 were presented at EUROSLA 24. Title: Incidental learning and online processing of L2 case-marking: A self-paced reading study. York: September 2nd, 2014.

6. Data from Experiment 3 were presented at AAAL 2016. Title: Working Memory, Awareness, and the Development of Implicit and Explicit Knowledge. Orlando: April 12th, 2016.

ACKNOWLEDGEMENTS

Undertaking a PhD is frequently likened to long and lonely endeavors, such as running a marathon or even a religious pilgrimage. I did not find this to be the case. It took a village for me to finish this thesis, and I would not have reached this milestone without the continued support and encouragement of many people along the way.

First, I have been truly blessed to have Andrea Révész as my supervisor. Words fail me in expressing my gratitude for her unwavering enthusiasm, promotion, and support throughout and beyond this project. I would not be where I am as a researcher without her guidance over the past four years. Andrea, I really cannot thank you enough.

I would also like to extend an explicit (!) note of gratitude to Patrick Rebuschat. The studies contained in this thesis took shape as a direct result of the groundwork laid by his previous research and were greatly improved as a result of his guidance and advice at the beginning stages of this project.

I would also like to acknowledge a few of my “SLA heroes”, specifically Ron Leow, Peter Robinson, and Marianne Gullberg. At different conferences over the past few years, these three have genuinely engaged with myself and other PhD students –

asking challenging questions and offering support and advice on how our work might move forward. This meant a lot to me as a PhD student, and serves as an example of the type of academic that I aspire to become.

I would like to thank the journal Language Learning, for supporting this thesis with a dissertation grant, the audience members at various conferences where aspects of this thesis have been presented, as well as the anonymous reviewers of journals where I have submitted parts of this work. This thesis was greatly improved as a result of their feedback, suggestions, encouragement and support. A note of gratitude is also due to the Foundation Program Department of English at Qatar University for allowing me the time and space needed to complete this thesis, as well as for the funding allowing me to present aspects of this research at conferences.

I would like to thank my parents for the unconditional support as well as the many, many sacrifices that they have made for us over the years. Thanks is also due to Sarah and Jacob, who, through our long history of competition, have indirectly motivated me to finish this thing. Finally, to my chief adviser, wife, best friend, and current boss, Lucie, your patience, understanding, and companionship have seen me through this project. I hope I haven't gotten too weird.

Collectively, I acknowledge from all parties a long narrative of support, encouragement, and belief that I was going to succeed. I pray this thesis does justice.

Table of Contents

List of Tables

List of Figures

For Lucie, who saw me through this project,
and Juliana, who joined us at the end.

When you are old and grey and full of sleep,
and nodding by the fire, take down this book,
and slowly read, and dream of the soft look
your eyes had once, and of their shadows deep

~William Butler Yeats~

CHAPTER 1

INTRODUCTION AND KEY CONCEPTS

Since the 1960s, much research within the cognitive sciences has focused on the role of conscious and unconscious processes in learning. This research has encompassed a range of research traditions, including artificial grammar learning (AGL) (Reber, 1967; see Reber, 1993 for a review), serial reaction time (SRT) research (Nissen & Bullemer, 1987; see Robertson, 2007 for a review) and the statistical learning paradigm, which has examined phenomena related to the learning of one's first language (L1) (Saffran, Aslin, & Newport, 1996; see Gómez, 2007 for a review). Studies operating within these areas have examined the learning of a wide variety of stimulus materials, both arbitrary in the form of artificial grammars, and stimuli that are more reflective of naturally occurring languages. This research has contributed to our understanding of the mechanisms of learning, as well as the constraints thereof. For instance, studies following the artificial grammar learning paradigm have been remarkably consistent in indicating that mere exposure to an artificial grammar results in a knowledge base that is largely tacit in nature (Reber, 1993; cf. Perruchet, 2008). Thus, this research has provided some evidence regarding the degree to which adults are able to develop knowledge of the grammatical regularities and patterns subsumed within the underlying structure of input without conscious awareness of having done so.[1]

[1] However, the exact nature of the knowledge that is acquired as a result of exposure to input in AGL studies has been a matter of fierce debate. See Perruchet (2008) and Section 3.1 for discussions.

Within the field of second language acquisition (SLA), interest in conscious and unconscious processes of learning came to the fore following several theories set down by Stephen Krashen (1982). Among Krashen's claims were arguments for two disparate learning processes in language learning: a conscious process that results in a conscious knowledge base, and an unconscious process that results in an unconscious knowledge base. Furthermore, Krashen postulated that explicit language instruction and metalinguistic awareness play little to no role in the acquisition process. These ideas sparked a great deal of empirical research into the effects of language instruction and the role of awareness in SLA, the results of which largely ran counter to many of Krashen's arguments. As an example, a wealth of studies has consistently demonstrated that higher levels of conscious awareness facilitate the learning of a second language (L2) (Alanen, 1995; Brooks & Kempe, 2013; DeKeyser, 1995; Grey, Williams, & Rebuschat, 2014, 2015; Hama & Leow, 2010; Leow, 1997, 2000, 2001; Rebuschat & Williams, 2012; Rosa & Leow, 2004; Rosa & O'Neill, 1999; Robinson, 1997, 2002, 2005; Williams, 2005).

Although there is now broad agreement concerning the facilitating effects of increased levels of conscious awareness, questions regarding implicit and explicit processes and products remain thorny topics within the SLA literature (Hama & Leow, 2010; Leow, 2015a, 2015b; Leow & Hama, 2013; Leung & Williams, 2011, 2012, 2014; Schmidt, 2010; Truscott & Sharwood Smith, 2011; Williams, 2005). Within this debate, recent years have seen a shift away from examining the role of awareness in the learning process towards gauging the conscious status of knowledge acquired as result of different learning conditions (see Rebuschat, 2013; Rebuschat &

Williams, 2012 for examples). Work in this area has included theoretical elaborations on the defining characteristics of explicit versus implicit knowledge (R. Ellis, 2004, 2005), as well as the development and refinement of a number of instruments designed to determine whether knowledge is implicit or explicit in nature (R. Ellis et al., 2009; Rebuschat, 2013; see Section 2.3 for a discussion of the characteristics of implicit versus explicit knowledge, and Section 3.5 for a discussion of the measurement of implicit versus explicit knowledge).

In summary, despite the fact that it is now well established that awareness and explicit instruction facilitate adult SLA, interest in topics related to implicit and explicit learning and the development of implicit and explicit knowledge continues to the present day. This interest has been sustained for a number of reasons. Firstly, on a theoretical level, the development and measurement of implicit and explicit knowledge is important because understanding the manner in which knowledge develops could shed light on how various aspects of languages are learned, how this knowledge is represented within memory systems, and how this knowledge manifests in L2 performance. Secondly, with regard to the role of awareness in the learning process, as noted above, it is now well established that awareness is facilitative to the learning process. However, it is also generally accepted that most language learning, both L1 and L2, takes place incidentally (Mitchell, Myles, & Marsden, 2013; VanPatten & Williams, 2007, 2015); in other words, without intention and as a result of processing input for meaning. Despite this, research has pointed to limits regarding the aspects of language learners appear to acquire on their own as a result of incidental exposure (Long, 2015; Schmidt, 2010; VanPatten, 2004a). As such, research in this area might serve to elucidate the constraints concerning what might be learned as a result of exposure to input, and to identify areas in which instructional intervention might be required.

The objective of the present thesis is two-fold. Firstly, this thesis sets out to examine whether L2 case markings, an area of grammar that is notoriously difficult for L2 learners to acquire (Larsen-Freeman, 2010; Long, 2015), can be learned under incidental learning conditions and to uncover the role of awareness in this process. For this purpose, a novel, semi-artificial language based on Czech was developed (see Section 1.3 for a brief overview of Czech case marking and Chapter 3 for a full discussion of the stimuli used in this thesis). Secondly, this thesis aims to reach a better understanding of the characteristics of the knowledge resulting from incidental exposure, in particular, to establish whether the resulting knowledge is implicit or explicit in nature. To do so, this thesis incorporates a number of methodological improvements to previous research. Specifically, this thesis incorporates a variety of instruments that are designed to differentiate between conscious and unconscious knowledge. By so doing, data concerning the conscious status of knowledge can be triangulated from multiple sources. This will allow for a more valid, and refined, view of the type of knowledge resulting from incidental exposure.

This thesis is organised as follows. Chapter 1 provides a brief introduction and the operationalisations of a number of key terms that are used extensively throughout this thesis as well as a brief introduction to Czech morphology. Chapters 2, 3, and 4 provide the background to the present investigation. Chapter 2 presents the theoretical underpinnings of this study. Specifically, it discusses the prevalent models of attention and awareness in SLA, sets out the defining characteristics of implicit and explicit knowledge, and provides an overview of the theoretical models that account for how implicit and explicit knowledge interface in their development. Chapter 3 has a methodological focus, and represents an overview of the two research paradigms that are germane to the present thesis: the artificial grammar learning paradigm and the serial reaction time paradigm. This chapter concludes with a critical evaluation of the instruments used to measure implicit and explicit knowledge within these paradigms. Chapter 4 then focuses on the literature relevant to the processing and acquisition of L2 case markers, and concludes with the research questions of the present thesis.

Chapters 5, 6 and 7 represent empirical chapters, in which the findings from a series of five experiments are presented. Chapter 5 reports on the results of two pilot studies conducted for the present thesis. The purpose of these pilot studies was, firstly, to test and validate the experimental materials and procedures and, secondly, to establish an initial learning effect among the participants. Chapter 6 examines the main experiments in this thesis, which followed the same research model as the previous two pilot studies. Chapter 6 consists of two experiments, Experiment 1 and its replication, Experiment 2. Chapter 7 then describes Experiment 3, which set out to address the same research questions as the previous experiments using a different research methodology, namely a self-paced reading task. Finally, Chapter 8 presents the conclusion of this thesis. It includes a discussion of the implications and limitations of the results of the five experiments discussed in this thesis, as well as possible directions for future research.

In the following section, several key items of terminology are discussed. These items are related to the role of consciousness in learning and are used extensively in this thesis. The following section is not intended to provide an in-depth discussion of these concepts, but rather to situate and operationalise these terms concisely and clearly as they are used throughout the present thesis. Beginning with clear definitions is particularly important when we take into account that differences in how researchers have operationalised many of these key concepts have contributed to the contentious nature of the ongoing debate on the role of consciousness in SLA (Dörnyei, 2009; Hulstijn, 2007; Leow, 2015a; Schmidt, 1995). I begin by briefly outlining previous attempts to arrive at a suitable 'blanket' definition for *consciousness* and its sundry sub-processes. Following this, further operationalisations are provided for the concepts that have dominated much of the debate within SLA, such as the distinction between *implicit* and *explicit* learning.

1.1. Key Concepts

1.1.1. Consciousness

There has long been a preoccupation with consciousness within the Western tradition (Baars, 1988). Attempts to come to an understanding of this term, as well as to flesh out what it means ‘to be’, can be found from the time of Plato’s Republic (c. 380 BC) to Descartes (c. 1640) and William James (1890), as well as many others in between. This interest continues to the present day, and has expanded beyond the borders of philosophy; it continues to be a current topic in a wide range of academic fields concerned with human cognition. As an illustration, several journals are focused directly on issues of consciousness within the field of cognitive psychology alone (such as *Consciousness and Cognition* and the *Journal of Consciousness Studies*). In addition, within this same field, recent years have seen a number of full-length manuscripts and reference books devoted to issues related to the relationship between human consciousness and cognition published (Bayne, Cleeremans, & Wilken, 2009; Blackmore, 2010; Velmans, 1996, 2009b; Velmans & Schneider, 2007; Zelazo, Moscovitch, & Thompson, 2007).

But what, exactly, does it mean to be conscious? The term consciousness, although seemingly a basic concept, has proved tricky to define. Within philosophy, proposed definitions have included associating consciousness with thought and the mind (for example, Descartes, 2010), or more recently, defining consciousness broadly in relation to subjective, personal experience (Nagel, 1974). Despite these attempts to define consciousness in global terms, there has been little agreement to date on the “core meaning” of consciousness. There does, however, appear to be consensus in that consciousness, and the concomitant features associated with consciousness, are a central aspect of being human (Velmans, 2009a, 2009b).

Although there have been a number of proposed definitions and theoretical accounts for consciousness, the most influential within cognitive psychology and SLA have been reductionist in nature. In other words, rather than attempting to broadly account for all aspects of human consciousness, these models have narrowed their focus to the relationship between consciousness and aspects of information processing (Baars, 1988, 1997a, 1997b; Block, 1996; 2007; Cleeremans, 2014; Cleeremans & Jiménez, 2002; Searle, 2007; Schacter, 1989). Among these models, Global Workspace Theory (Baars, 1988, 1997a, 1997b) has arguably been the most influential on SLA. In this model, consciousness is defined as being the "facility for accessing, disseminating and exchanging information, and for exercising global coordination and control" (Baars, 1997b, p. 19). Within Global Workspace Theory, consciousness can be seen as being a central aspect of both general cognitive processes and of language.

Despite adopting a reductionist view of consciousness, there is still considerable controversy and disagreement concerning how to define and operationalise this term within the field of SLA. This disagreement can partly be attributed to the fact that consciousness is not a unitary construct, but is comprised of several sub-processes, such as attention, awareness, intentionality, and knowledge (Cleeremans & Jimenez, 2002; Schmidt, 1990, 1994a; see Velmans, 2009b for a full discussion), each of which has also proven difficult to define and operationalise (see discussion below). In an attempt to move the debate forward, several researchers (Leow, 2002) have suggested dropping consciousness as a blanket term to cover all of these sub-processes. In this thesis, I follow this suggestion and refer to a specific sub-process, such as awareness, where relevant. Where the term consciousness is used, it is used in the everyday sense of being synonymous with awareness. Constructs related to the sub-processes of attention, awareness, intentionality, and knowledge are further discussed and operationalised below.

1.1.2. Intentional versus Incidental Learning

Concerning intentionality, the constructs of *incidental* and *intentional* learning have garnered considerable attention in SLA (Hulstijn, 2003). Firstly, intentional learning is learning that occurs with intent. An example of intentional learning in an instructed setting would be students trying to learn a set of descriptive L2 grammatical rules in a language classroom. In experimental contexts, intentional learning has also been operationalised as a “rule-search” training condition in which participants do not receive direct instruction, but are informed of the existence of a rule or rules and are given the task of discovering these rules during the exposure phase of the experiment (Rebuschat, 2008; Robinson, 1996). What these two examples have in common is that both the students receiving instruction in the classroom and the participants in the rule-search condition are aware that there are features that they should be learning and, furthermore, are typically given the task of trying to learn these features.

Less straightforward than *intentional learning* is the term *incidental learning*. A basic definition would be that incidental learning refers to learning that occurs without intent. A narrower definition of this term was proposed by Schmidt (1994a) as the learning "of one thing (e.g., grammar) when the learner's primary objective is to do something else (e.g., communicate)"(p. 16). An example of this might be someone who picks up a new word or phrase as a result of engaging in active communication (Hulstijn, 2013). From a methodological perspective, incidental learning has been operationalised as an experimental condition in which participants are not informed that they will be tested later in the experiment (Hulstijn, 2003, 2013; Williams, 2009). Frequently, experiments operating in this area also include a cover story, or task in their training conditions, which is designed to orient participants' attention towards the meaning of the input, rather than towards the grammatical features that will later be tested (see, for example, the plausibility judgement task in

Rebuschat & Williams, 2012, or the discussion of supraliminal tasks in Shanks & St. John, 1994).

It is important to stress, however, that incidental learning conditions do not control or guarantee that learners will necessarily learn incidentally (Schmidt, 1993). Several studies have demonstrated that some learners may employ intentional learning strategies even under incidental learning conditions (Alanen, 1995; DeKeyser, 1995; Leow, 1998, 2000; Robinson, 1996, 1997). As such, incidental learning is operationalised here in a strictly methodological sense—specifically, as an experimental condition designed to influence the manner in which participants approach the task(s), and consequently process the input, in the training phase of an experiment.

1.1.3. Attention and Awareness

The most influential model of attention and awareness in SLA stems from Schmidt's noticing hypothesis (1990, 1993, 1994a, 1995, 2001, 2010). In this model, Schmidt differentiates between two levels of awareness that are argued to be facilitative for the learning process (Schmidt, 1994a). Firstly, *awareness at the level of noticing* (henceforth *noticing*) refers to a lower level of awareness that involves the conscious registration of surface features of the input. For Schmidt (1995), this level of awareness is largely isomorphic with focal attention. Next, reflecting a higher level of awareness, *awareness at the level of understanding* reflects awareness that includes rule and/or metalinguistic knowledge. Schmidt's noticing hypothesis, together with other influential models of attention and awareness in SLA (Robinson, 1995; Tomlin & Villa, 1994) will be discussed further in Chapter 2 (see Section 2.2.1).

1.1.4. Implicit versus Explicit Learning

Two types of learning are directly related to the issue of awareness in SLA: implicit and explicit learning. The term *implicit learning* is often conflated with incidental learning (Hulstijn, 2007; Leow, 2015a). This confusion can be partly attributed to the fact that implicit learning is learning that also takes place incidentally; that is, without intention. However, implicit learning is further characterised by a lack of awareness throughout the learning process (Williams, 2009). From a methodological perspective, when learning incidentally, participants may or may not become aware of the linguistic focus of the experiment. The key criterion when learning implicitly, on the other hand, is that participants remain unaware of what is being learned during the experiment. Explicit learning is often equated with intentional learning in that it refers to learning that occurs with intent and conscious awareness of what is being learned (DeKeyser, 2003). An example of explicit learning, as with the example of intentional learning above, would be scenarios in which people are deliberately trying to learn; for example, students studying grammatical rules on which they will later be assessed at school.

1.1.5. Implicit versus Explicit Knowledge

In contrast to implicit and explicit learning, which refer to the learning process, the terms implicit and explicit knowledge refer to the product of learning (Leow, 2015a, 2015b). Implicit knowledge is unconscious knowledge–in other words, implicit knowledge lies outside of awareness, cannot be verbalised and, as such, can only be inferred from behaviour (R. Ellis, 2004, 2005; Loewen, 2015; Reber, 1993). Explicit knowledge, then, is knowledge that we are aware we possess, that lies within awareness and is often, though not always, verbalisable (DeKeyser, 2009; Dörnyei, 2009; Reber, 1993; Williams, 2009). As noted above, although there is considerable debate concerning what it means to be conscious (see Velmans, 2009b), following DeKeyser (2009), in this thesis I treat the terms *implicit* as being synonymous with *unconscious*, and *explicit* as being synonymous with *conscious*. Further distinctions between implicit and explicit knowledge, as well as their interface, are discussed in Chapter 2 (see Sections 2.4-2.5).

1.1.6. Declarative versus Procedural Knowledge

Although often associated with skill-acquisition theory (Anderson, 1982; Newell & Rosenbloom, 1981), declarative and procedural knowledge are also important concepts that feature heavily in the debate on consciousness in SLA (R. Ellis, 2004, 2005). *Declarative knowledge* is typically defined as "knowledge that," whereas *procedural knowledge* is "knowledge how" (DeKeyser, 2009). What this means is that declarative knowledge is factual knowledge, such as knowing how many degrees there are in a right angle, or being able to describe how one rides a bicycle. Procedural knowledge, on the other hand, could represent motor skills, such as how to actually ride a bicycle in practice, or mental skills, such as how to solve geometrical problems (DeKeyser, 2001; 2007a, 2007b, 2007c; 2009). Although declarative/procedural knowledge is often confused with explicit/implicit knowledge, these terms are not isomorphic (DeKeyser, 2009; DeKeyser & Juffs, 2005; Reber, Allen, & Reber, 1999; Reber & Squire, 1999; Squire & Knowlton, 2000; Ullman, 2005; Williams, 2009). A key difference is that while explicit / implicit knowledge is considered to be represented within distinct memory systems, which may or may not influence and interact with each other,[2] declarative and procedural knowledge are seen as part of a continuum of skill development in which declarative knowledge can become functionally equivalent to procedural knowledge as a result of time and practice.

[2] This, of course, depends on one's theoretical orientation regarding the interface between implicit and explicit knowledge. As noted above, this topic is discussed in more detail in Section 2.4 below.

CHAPTER 2

THEORETICAL FOUNDATIONS

As mentioned in the introduction, Chapter 2 of this thesis provides the theoretical background to the present thesis, particularly with regard to issues surrounding the role of awareness in SLA, and the debate concerning the development of implicit and explicit knowledge. This chapter first briefly introduces Krashen's (1982) influential monitor theory before discussing more current models that address the role of attention and awareness in SLA, specifically Schmidt's (1990) noticing hypothesis, and Tomlin and Villa's (1994) model of attention in SLA. Following this, the focus of this chapter shifts to how implicit and explicit knowledge have been classified and defined in the literature. The chapter concludes with an overview of the various interface positions that have attempted to account for the manner in which implicit knowledge and explicit knowledge interact in their development.

Implicit learning, or learning that occurs without intention and in the absence of awareness of what has been learned (Perruchet, 2008; Reber, 1993; Shanks & St. John, 1994; Williams, 2009), has been argued to bear on many areas of everyday life. For instance, the mastery of social skills is generally acknowledged to develop implicitly (Perruchet, 2008; Reber, 1993). There also is substantial experimental evidence confirming that people can acquire unconscious knowledge of complex artificial grammars solely as a result of exposure to input (see Cleeremans, Destrebecqz, & Boyer, 1998 and Reber 1989, 1993, for reviews). With regard to language learning, first language (L1) acquisition is frequently cited as a process that occurs implicitly (N. Ellis, 2002, 2008, 2011; Long, 2007; Reber, 1967; Van den Branden, 2007; Wells, 1985).[3] The degree to which implicit processes are involved in second language (L2) learning has, however, been the focal point of much debate within the field of second language acquisition (SLA) in recent years (Andringa & Rebuschat, 2015; Chaudron, 1983; DeKeyser, 1997, 2003, 2009; Dörnyei, 2009; Dörnyei & Ryan, 2015; N. Ellis, 2005, 2011, 2015; R. Ellis, 1994, 2015; Gass, 1997; Han & Finneran, 2014; Hama & Leow, 2010; Hulstijn, 2015; Krashen, 1978, 1981, 1982, 1985; Leow, 2002, 2013, 2015a, 2015b; Leow & Hama, 2013; Leung & Williams, 2011, 2012, 2014; Long, 1988, 1996, 2015; McLaughlin, 1990a; Paradis, 2004, 2009; Rebuschat, 2015a; Robinson, 1995; Robinson, Mackey, Gass, & Schmidt, 2012; Schmidt, 1990, 1993, 1994a, 1994b; 1995, 2001, 2010; Sharwood Smith, 2008; Truscott, 1998, 2015a, 2015b; Truscott & Sharwood Smith, 2011; Williams, 2005; see DeKeyser, 2003 & Williams, 2009, for reviews). This interest was largely sparked by the work of Stephen Krashen (1981, 1982, 1985). Among other claims, Krashen has argued that both L1 and L2 acquisition are primarily the result of an unconscious (implicit) process, and occur as a result of processing input for meaning.

2.1. Krashen and Monitor Theory

[3] Although L1 acquisition is generally considered the result of an unconscious process, some have expressed quiet reservations about the empirical support underlying this assumption. For instance, Long (2007), in a discussion of monitor theory, notes that "children (supposedly) learn their first language without conscious awareness" (p. 17).

Building on earlier work by Corder (1967), Krashen (1978, 1981, 1982, 1985) distinguished between the types of knowledge acquired from two distinct processes, which he labelled *learning* and *acquisition.*[4] *Learning*, in Krashen's view, is a conscious process characterised by learning *about* language and its structure (i.e., metalinguistic information). One might associate *learning* with what students learn within a more traditional, non-communicative, language classroom; for example, memorising prescriptive rules for the use of verb tenses in a foreign language. In contrast, *acquisition* is a process similar to first language acquisition in that it comes about as a result of exposure to input. Furthermore, and importantly, *acquisition* is an unconscious process (1982). This distinction between *learning* and *acquisition* is a central part of Krashen's monitor theory. In this theory, only *acquisition* can lead to communicative competence—in other words, the ability to use language spontaneously, freely, and fluently. The knowledge that results from *learning*, on the other hand, is knowledge *about* language (1985). In this view, knowledge about language is not directly related to communication: rather, it can only serve to monitor the accuracy of the language being produced. In sum, for Krashen, explicit language learning (as it might be referred to today) and metalinguistic awareness play little or no role in either the L1 or the L2 acquisition process. To be able to use a language, one must acquire it unconsciously, over time, and as a result of exposure to input.

Krashen's monitor theory encountered much criticism (Gregg, 1984; Hulstijn & Hulstijn, 1984; Sharwood Smith, 1981), and it has been superseded by alternative

[4] Although Krashen distinguishes between learning and acquisition, these two terms are used interchangeably throughout this thesis to refer to the development of L2 knowledge. Throughout this thesis, I use the italicised forms *learning* and *acquisition* to indicate Krashen's usage.

theories regarding the role of consciousness in SLA (Loewen, 2015). These include alternative models of attention and awareness in SLA, which are discussed in detail in the section below. Despite falling out of favour with present-day SLA researchers, Krashen's influence on the field is undeniable; his arguments about *learning* versus *acquisition* brought issues of attention, awareness, and types of L2 knowledge to the fore in SLA research. I return to some of Krashen's arguments in Section 2.4, in which I discuss the interface between implicit and explicit knowledge.

2.2. Attention and Awareness in SLA

2.2.1. Schmidt's Noticing Hypothesis

Running counter to previous theories (such as that of Krashen, 1981) that posit a minor role for conscious processes in SLA, Schmidt (1990, 1993, 1994a, 1994b, 1995, 2001, 2010) called for a re-evaluation of the role of consciousness in language learning. Building upon case studies of second language learners, including his own experience of learning Portuguese (Schmidt, 1983; Schmidt & Frota, 1986), Schmidt argued that awareness is facilitative, and perhaps even necessary, in second language acquisition. In what is now known as the noticing hypothesis, Schmidt drew on a wide range of research from the cognitive sciences to develop a model to account for the role of awareness in SLA. There are three levels in Schmidt's model of awareness. The first level, *perception*, is unconscious: input that is only registered at this level does so without any conscious awareness of the input being processed. Schmidt pointed out that while subtle effects on behaviour can be seen from information processed below the threshold of awareness, it is unclear whether these effects should actually be considered as learning. Furthermore, there is an absence of “indisputable evidence about events near the threshold of conscious experience” (1993, p. 210). It should be pointed out that similar scepticism regarding the possibility of subliminal learning is not limited to Schmidt's noticing hypothesis: Whether or not learning can occur below the threshold of awareness remains a topic of debate in, for example, cognitive psychology (Perruchet, 2008; Shanks & St. John, 1994).

Noticing (awareness at the level of noticing) is the second level in Schmidt's model. For Schmidt (1995), noticing is isomorphic with focal attention. It can be distinguished from perception in that it is marked by conscious awareness of, at least, the surface features of the input. Noticing has been operationalised as the “conscious registration of the occurrence of some event” (1995, p. 29) and is “related to rehearsal within working memory and the transfer of information to long-term memory, to intake, and to item learning” (Schmidt, 1993, p. 213). Noticing, then, represents awareness at a low level of abstraction, in that the objects of noticing are the surface features of the input, rather than underlying grammatical rules or patterns (Schmidt, 2001).

The final level in Schmidt’s model is *understanding* (awareness at the level of understanding). While awareness at the level of noticing may fit the common definition of what constitutes being "aware" of something, awareness at the level of understanding reflects further processes that occur within consciousness to attempt to explain or make sense of observed phenomena. Schmidt notes that both problem-solving and all forms of metacognition are indicative of awareness at this level (Schmidt, 1990) and that awareness at this level “implies recognition of a general principle, rule or pattern” (Schmidt, 1995, p. 29). Furthermore, if noticing is related to the transfer of information to long-term memory and item learning, then awareness at the level of understanding is connected to “the organization of material in long term memory, to restructuring, and to system learning” (Schmidt, 1993, p. 213).

In what can be referred to as the strong form of the noticing hypothesis, Schmidt argued that awareness at the level of noticing is a necessary and sufficient condition for SLA to occur. Understanding, while perhaps not necessary, is argued to have a facilitating effect on SLA. This strong form of the noticing hypothesis has, however, encountered criticism in relation to a number of theoretical and methodological issues, as well as in the face of contradictory empirical findings. For instance, Schmidt (2001) has acknowledged that his original claim that no learning can occur without awareness is not falsifiable and thus cannot be tested empirically (Baars, 1997a; Dienes, 2008a; McLaughlin, 1990a). In addition, in opposition to Schmidt's claim that attention to input is a necessary condition for SLA, Gass (1997) has argued that not all learning appears to depend on input. Studies that have examined the universal relative clause accessibility hierarchy (Eckman, Bell, & Nelson, 1988; Gass, 1979, 1982) have indicated that, after receiving instruction on particular relative clause types (such as the object of a preposition), learners were able to generalise this knowledge to completely different relative clause forms (such as subject or object relative clauses) to which they had not been exposed in the input. This finding casts some doubt on the strong version of the noticing hypothesis because these data indicate learning in the absence of input. Thus, "if no input [to the target structures] existed, how could attention to input be a necessary condition for all aspects of learning?" (Gass, 1997, p. 16).

Given these criticisms, the strong form of the noticing hypothesis, as Schmidt has conceded, makes too bold a claim (Schmidt, 1995, 2001; Truscott, 1998). Schmidt later modified the noticing hypothesis so that awareness at the levels of both noticing and understanding are facilitative to SLA (Schmidt, 1994b; 1995; 2010), with noticing perhaps still being necessary under certain conditions; for example, with non-salient or redundant grammatical features (2001). It is important to reiterate that these arguments against the strong form of the noticing hypothesis should not be interpreted as casting doubt on the importance of attention and awareness in the process of learning a second language. Schmidt (2001, and onwards) has maintained his position regarding the importance of attentional processes in SLA, a view that is also subsumed within a number of theoretical frameworks in SLA (N. Ellis, 2005; Gass, 1997; Leow, 2015a, 2015b; Robinson, 1995) (this point is touched on further at the end of Section 2.2.2).

2.2.2. Tomlin and Villa's Model of Attention

Although Schmidt (1995) considered awareness at the level of noticing to be isomorphic with focal attention, Tomlin and Villa (1994) proposed a model of attention that, they argued, might lead to a better understanding of the attentional processes involved in SLA. The central argument of Tomlin and Villa (1994) is that previous theoretical formulations of attention in SLA were influenced by four different conceptualisations of attention:

a) The limited capacity model (Anderson, 1983, 1990),

b) selective attention (Wickens, 1989),

c) controlled versus automatic processing (McLaughlin, 1990b), and

d) conscious regulation of attentional resources (Bialystok, 1990, 1992).

These conceptualisations, Tomlin and Villa argued, are too "coarse-grained" in that they do not provide specific information as to how attention is allotted during a task. Drawing on the work of Posner (1988, 1990; 1992; Posner & Carr, 1992), Tomlin and Villa proposed a finer-grained model in which attention is not analogous to awareness (cf. Schmidt, 1990), and attention is not a unitary construct, but is instead comprised of three separate but interrelated processes.

Alertness, the first level in Tomlin and Villa's (1994) model, describes an overall mental readiness to deal with incoming information. Tomlin and Villa argued that it has been shown that alertness can be both assessed and manipulated under experimental conditions and that different levels of alertness have “differential effects on processing or performing tasks” (Tomlin & Villa, 1994, p. 190). In short, alertness is important to SLA because input is more likely to be subjected to further processing when learners are mentally prepared to deal with input.

Orientation, the second level in this model, concerns the outcome of the allocation of attentional resources to stimulus material. It is at this stage that the learner “aligns” or “orients” his or her attention to a type or class of input (such as auditory information) at the expense of others (for example, visual information). This process has been argued to involve higher-level activation of schema that “may predispose learners to attend … to form or meaning in processing a stimulus” (Robinson, 1995, p. 296). Tomlin and Villa noted that input that the learner is oriented towards has a greater chance of reaching detection, the next stage of attention. This does not preclude the possibility of stimuli that occur outside of orientation reaching the stage of detection: Stimuli that do not receive attentional orientation may still be detected, but the possibility of this occurring is inhibited and would require a substantial amount of effort on the part of the learner.

Detection is the final stage of Tomlin and Villa’s model. Unlike orientation, in which a learner aligns his or her attention with a particular type or class of stimuli, detection is the process “that selects, or engages, a particular and specific bit of information” (Tomlin & Villa, 1994, p. 192). Detection consumes additional attentional resources and enables “further processing of a stimulus…such as storage and rehearsal in short-term memory” (Robinson, 1995, p. 296). In summary, detection represents the cognitive registration of stimuli and, as Tomlin and Villa argued, it is at the level of detection that language acquisition must take place. Furthermore, it is detection alone that is necessary for the further processing of incoming stimuli. The levels of alertness and orientation can increase the probability that detection will take place, but neither of these two levels is necessary.

As previously mentioned, Schmidt (1995) considered noticing to be isomorphic with focal attention. As such, it would appear that detection, in Tomlin and Villa's model, can be seen as roughly analogous to awareness at the level of noticing in Schmidt’s model (1990, 1995, 2001). It is important to note that, unlike Schmidt's noticing hypothesis, none of the levels in Tomlin and Villa's model, including detection, requires awareness. Similarly, Robinson (1995) has pointed out that there is “considerable evidence indicating that information can be cognitively detected, even though the individual is not aware its having occurred" (p. 193). This represents the central difference between Schmidt's noticing hypothesis and Tomlin and Villa's model. In the latter model, learning can take place outside of conscious awareness, whereas the strong version of Schmidt's noticing hypothesis does not allow for this possibility. In order to reconcile these differences, Robinson (1995) proposed redefining the concept of noticing to "detection plus rehearsal in short-term memory, prior to encoding in long-term memory" (p. 296). Noticing can thus refer to exemplars that are detected in Tomlin and Villa's model, and can then be "further activated following the allocation of attentional resources from a central executive" (Robinson, 1995, p. 297). In this view, awareness remains the defining characteristic of noticing and, furthermore, clearly delineates the constructs of noticing and detection.

2.2.3. Summary

Schmidt's noticing hypothesis and Tomlin and Villa's model of attention are concerned with the role of attention and awareness in the learning process. For Schmidt (1990, and onwards), attention is considered identical to awareness in the form of noticing, and awareness at the level of noticing plays a facilitative and, at times, vital role in SLA. By contrast, in Tomlin and Villa's model, detection is considered sufficient for learning to take place, and detection can occur on both a conscious and on an unconscious level. In other words, awareness is divorced from attention, and awareness is not necessary for SLA to take place. Despite various attempts to reconcile the differences between these two models in a single framework (Robinson, 1995; Schmidt, 2001; Truscott, 2015a, 2015b; Truscott & Sharwood

Smith, 2011), the role of consciousness in SLA remains a current topic of interest within SLA (see Andringa & Rebuschat, 2015; Rebuschat, 2015a; Williams, 2012). In addition to examining the role of attention and awareness at the point of learning, recent years have seen an increased focus on the conscious status of linguistic knowledge; in other words, whether L2 knowledge gained through various interventions is implicit or explicit in nature. The next section addresses an area central to this thorny area of enquiry: the differentiating characteristics of implicit versus explicit knowledge, and the operationalisation of these two terms within SLA.

2.3. Nature of Implicit and Explicit Knowledge

As noted in Section 2.2, Krashen (1981, 1982, 2003) distinguished between two independent learning processes; *learning*, which is a conscious process, and unconscious *acquisition*. Furthermore, as part of this argument, Krashen maintained that the conscious process of *learning* results in a conscious knowledge base, namely explicit knowledge. According to Krashen, one example of this conscious knowledge is metalinguistic knowledge, or knowledge about the rules of language. Importantly, for Krashen, the conscious knowledge resulting from *learning* is not available for spontaneous communication: It can only serve as a monitor of language being produced. In contrast to *learning*, *acquisition* is an unconscious process that comes about solely as a result of exposure to input. Furthermore, *acquisition* results in an unconscious knowledge base of implicit knowledge that can be used for spontaneous, free language production (Krashen, 1982).

As noted in Section 2.1, much of Krashen's work, such as his distinction between *learning* and *acquisition*, no longer represents received views within the field of SLA. However, some of Krashen's arguments do remain topics of debate. For example, his claims for a relationship (or lack thereof) between conscious and unconscious knowledge continue to garner interest within SLA. This interest has mainly taken the form of theoretical accounts that attempt to map the respective roles of implicit and explicit learning in second language development (see N. Ellis, 2005; R. Ellis, 1994; Han & Finneran, 2014; Paradis, 2009).

Of particular importance to this debate is the definition and operationalisation of implicit and explicit knowledge, and how these two types of knowledge are manifested within linguistic performance. As noted in Section 1.2, implicit and explicit knowledge are often operationalised and defined, as is the case in this thesis, as being unconscious/conscious types of knowledge, respectively. Recent years have, however, seen attempts to elaborate on the operational characteristics of implicit versus explicit knowledge (DeKeyser, 2009; R. Ellis, 2004, 2005; R. Ellis et al., 2009; Reber, 1993). For example, it has been argued that implicit and explicit knowledge also differ in their degree of verbalisability, learnability, and automaticity. In addition, it has been postulated that implicit and explicit knowledge differ with regard to how they are utilised, depending on the nature of the task at hand. In the next section, I will elaborate further on these contrasting features of implicit and explicit knowledge.

2.3.1. Consciousness

As mentioned previously, implicit knowledge is tacit knowledge (R. Ellis et al., 2009; Reber, 1993; Rebuschat, 2013; Reingold & Merikle, 1988, 1990; Shanks & St. John, 1994). In other words, it represents knowledge that we are not aware that we possess. Conversely, knowledge can be considered explicit if we are aware that we possess it. During performance, implicit knowledge is commonly realised through feelings of intuition (Dienes, 2008b, 2010; Dienes & Scott, 2005). Explicit knowledge, on the other hand, is often evinced through the use of conscious knowledge of, for example, an underlying grammatical system. This conscious knowledge can be partial, in the form of bigrams and trigrams of an artificial grammar system (Perruchet, 2008; see Section 3.1 for a brief discussion), or complete in the form of abstract rule knowledge of the underlying structure or principles governing a linguistic system (see Pothos, 2007; Reber, 1967, 1969, 1976; Shanks, 2005). A classic illustration of this distinction is how native speakers might judge the grammaticality of a sentence (R. Ellis et al., 2009). When confronted with the ungrammatical sentence *He has gone to the store last week,* a native speaker of English would know that something is wrong with it. If you asked this native speaker how he or she knows that this sentence is ungrammatical, the expected response would be that it just sounds, or feels, incorrect. In other words, a native speaker would rely on implicit knowledge, realised in the feeling of intuition, to make this judgement. Conversely, a native speaker with more metalinguistic knowledge, such as linguists or EFL teachers, would know that the present perfect construction typically refers to events that occurred at an unspecified time in the past. In this case, the knowledge here would be conscious in that the native speaker is aware that he or she knows the rule that he or she is using to judge the sentence's grammaticality.[5]

[5] To cloud this issue further, however, this native speaker would be in possession of both implicit knowledge as a result of learning the L1, and explicit knowledge as a result of studying grammatical rules. It is possible that a native speaker in this situation could rely on his or her implicit knowledge to judge the sentence, and later use his explicit, metalinguistic knowledge to justify his or her previous

2.3.2. Verbalisability

Closely related to consciousness, verbalisability has historically been the operationalisation most commonly used to determine whether knowledge is implicit or explicit in nature (Baars, 1988; Rebuschat, 2013; Seth, Baars, & Edelman, 2005). If knowledge is implicit, then it is not verbalisable. As in the example in Section 2.3.2, a native speaker without formal linguistic training would base his or her decision on feeling and would not be able to explain the grammatical rule shepherding his or her decision. Implicit knowledge, then, can only be inferred from behaviour—in this case, by observing the native speaker's performance, we can infer that some knowledge is driving his or her (accurate) decisions. By logical extension, then, explicit knowledge is verbalisable to some degree, whether in the form of a complete rule or a partial rule, and can be expressed in technical or layman's terms (Loewen, 2015). Despite its long use as a criterion for implicitness in the cognitive sciences, verbalisability has been criticised on methodological, although not necessarily theoretical, grounds (Leow, 2002; Reingold & Merikle, 1988; Schmidt, 1994a; Shanks & St. John, 1994). I will return to these criticisms related to the measurement of implicit and explicit knowledge in Section 3.5.

2.3.3. Automaticity

Another operational distinction between implicit and explicit knowledge lies in the relative automaticity of these two types of knowledge (DeKeyser, 2009). Explicit knowledge has been argued to consist of facts, or rules, such as metalinguistic knowledge of grammatical rule systems. Furthermore, explicit

(implicit) judgment (see Section 2.3.6 below for a brief discussion of how task demands influence the type of knowledge utilised by the learner).

knowledge is available for controlled processing; that is, explicit knowledge can be accessed less quickly than is implicit knowledge, and, it is argued can only be utilised when sufficient time is provided to access it. Conversely, implicit knowledge has been argued to be available for automatic, spontaneous processing (Bialystok, 1982; R. Ellis, 2004, 2005). In other words, parallels have been drawn between explicit and declarative knowledge, as well as between implicit and procedural knowledge. However, implicit/explicit knowledge is not synonymous with procedural/declarative knowledge. These terms have developed via different theoretical frameworks (this point is discussed further in Section 2.4). For example, declarative/procedural knowledge is thought to operate along a continuum on which declarative knowledge can be transformed into procedural knowledge as a result of practice (Anderson, 1982, 1992, 2000; Anderson & Schunn, 2000; Carlson, 2003; DeKeyser, 1997; 1998, 2007b; R. Ellis, 1993). Implicit and explicit knowledge, on the other hand, are generally regarded as distinct systems, the interface of which is a matter of much debate in SLA (see Section 2.4, and Robinson et al. 2012, for a concise review).

2.3.4. Learnability

Still another operational difference between implicit and explicit knowledge has been argued to be the degree to which these various types of knowledge develop in adult learners of an L2. Here, learnability should not be interpreted following Krashen's distinction between *learning* and *acquisition* (see Section 1.2). To say that explicit knowledge is more learnable than implicit knowledge means that there appear to be fewer limitations to the development of explicit knowledge compared to the development of implicit knowledge. There are two issues at play with regard to the learnability of implicit versus explicit knowledge. The first concerns limits on the acquisition of implicit and explicit knowledge in light of individual difference variables such as age and aptitude. This argument has largely arisen out of the putative constraints on the development of implicit knowledge when taking into account that many L2 learners, despite years of formal instruction, fail to approach a near-native level of proficiency in their L2 (Loewen, 2015, Long, 2013, 2015). Other researchers (Bialystok, 1982; DeKeyser, 2003; N. Ellis, 2005) have argued that explicit knowledge can be learned throughout one's lifespan. In this view, older learners are more adept at learning explicitly (Muñoz, 2006, 2008), whereas there are age-related limitations to the development of implicit knowledge by L2 learners (Long, 2015).

A further issue with regard to differences in learnability between implicit and explicit knowledge concerns limitations concerning the linguistic features that can be learned implicitly versus explicitly. Both Reber (1993) and Krashen (1981, 1982) have theorised that only relatively simple rules, such as third person *–s,* can be learned and formalised as explicit knowledge. By contrast, they are that only complex rules can only be "acquired", or developed, through an unconscious process that results in implicit knowledge. Although few studies in SLA have tested these claims directly (DeKeyser, 1994, 1995; Robinson, 1996), what little evidence there is suggests that explicit learning is best suited to rules that are simple and categorical, whereas implicit learning is more suited to concrete elements in close proximity, such as collocations, or to more complex or non-salient rules (Long, 2015; see also N. Ellis, 2005 for a discussion).[6]

2.3.5. Task demands

Finally, with regard to language use, implicit and explicit knowledge can be utilised to varying degrees depending on the nature of the task. Evidence from Bialystok (1982) suggests that the demands of task type might influence the types of knowledge deployed in terms of the degree of analysis and control. For example, error correction tasks presented visually encourage learners to rely on explicit knowledge. The same task presented auditorily, however, promotes the use of implicit knowledge. In addition, tasks that the L2 learner finds difficult, or that result in breakdowns in communication (N. Ellis, 2005), might result in the learner attempting to exploit explicit knowledge. R. Ellis (2004) gives the example of learners being asked to justify their choices concerning a GJT on a think-aloud task, during which they might attempt to draw on their explicit knowledge if they do not feel confident enough to make a judgement based on their intuition. Such tasks have been operationalised as being likely to elicit more implicit or explicit knowledge according to their modality (visual versus auditory), as well as with regard to whether they encourage automatic versus controlled processing (see R. Ellis, 2005; Rebuschat & Williams, 2012). Such operationalisations, however, should not be interpreted as suggesting that the use of implicit and the use of explicit knowledge are mutually exclusive. It is generally well accepted (DeKeyser, 2009; N. Ellis, 2005; R. Ellis,

[6] However, as others have pointed out (Erlam, 2013; Roehr, 2008, 2010), it is not always straightforward to identify exactly what constitutes a “simple” versus a “complex” rule.

2004; Loewen, 2015) that learners utilise both implicit and explicit knowledge stores to varying degrees depending on the nature of the task at hand.

2.3.6. Summary

In summary, explicit knowledge is marked by awareness—in other words, we know that we know it. Explicit knowledge is typically verbalisable, learnable, declarative, and not spontaneous, and it can be elicited by tasks that allow for controlled processing. Conversely, implicit knowledge is by definition unconscious. It is not verbalisable, is typically procedural in nature, can usually be deployed spontaneously and automatically, and can be elicited by tasks that encourage faster, more automatic processing. This section has outlined the definitions and operational characteristics of implicit and explicit knowledge. At this point, it is worth shifting our focus towards how these two types of knowledge develop and, furthermore, interact in their development.

2.4. Interface between Implicit and Explicit Knowledge

2.4.1. Non-interface

To return to Krashen (1981, 1982, 1985), the monitor theory posited that adult second language acquisition is fundamentally the same as first language acquisition by children. Furthermore, this process is both incidental, in that it happens without intention, and implicit, in that acquisition occurs without awareness and is entirely based on the abstraction of patterns from the input. Krashen’s monitor hypothesis claims that learners possess two distinct systems that are involved in language learning: a "learned system" and an "acquired system." As mentioned earlier, *acquisition* is an unconscious (implicit) process that occurs during communication, whereas *learning* is the conscious (explicit) process of attempting to understand language and/or memorise rules. For Krashen, unconscious acquisition is responsible for the vast majority of what is ultimately learned. *Learning*, on the other hand, plays a minor role in this process and only serves to monitor the output the learner produces. In other words, rules that are explicitly learned only serve to compare with, and possibly make changes to, utterances that have been produced by the "acquired system" (Krashen, 1982). Of particular relevance here, Krashen also claimed that the two different systems, learned and acquired, are not connected: Thus, learned competence cannot turn into acquired competence (1982, 1985). The position that explicit knowledge cannot become implicit knowledge later became known as the non-interface position (R. Ellis, 2008); it inspired a number of competing theories regarding the role of explicit knowledge in the development of implicit knowledge. Firstly, non-interface positions (Krashen, 1982, 1994; Paradis, 1994, 2004, 2009) hold that explicit knowledge cannot be transformed into implicit knowledge, and explicit knowledge does not directly influence the development of implicit knowledge. The strong-interface position (DeKeyser, 2003, 2007a, 2007b, 2007c, 2009, 2015), which is predicated on skill-acquisition theory (please see Section 2.4.2 below), argues for a direct line of causation from declarative knowledge to procedural knowledge. Finally, weak-interface positions (N. Ellis, 2005; R. Ellis, 1994) argue that explicit knowledge can directly facilitate the development of implicit knowledge, although the success of this hinges on various factors. These positions are discussed in more detail in the sections below.

2.4.2. Strong Interface

Early critics of Krashen's distinction between *learning* and *acquisition* suggested alternative accounts to explain the relevant phenomena cited in support of the monitor hypothesis, such as the relative speed of production using implicit knowledge compared to when using explicit knowledge (see Krashen, 1982, p. 89). Building on an information-processing model developed within the domain of cognitive psychology, McLaughlin (1978) proposed distinguishing between controlled and automatic processes. The central claim of this model, later to be identified as skill acquisition theory (see Anderson, 1982; Carlson, 2003; DeKeyser, 1998, 2007b for overviews), is that "a wide variety of skills shows a remarkable similarity in development from initial representations of knowledge…to eventual fluent, spontaneous, largely effortless, and highly skilled behaviour" (DeKeyser, 2007a, p. 97). This process is governed by the power law of practice (Anderson, 2000), which is a mathematical function that describes how error rates and reaction times decline over time "regardless of the domain of learning" (DeKeyser, 2007b, p. 3).

It is important to note that this position, which would later be called the interface hypothesis, or strong interface position (R. Ellis, 2008), does not typically refer to implicit versus explicit knowledge. Instead, the strong-interface position references declarative and procedural knowledge. These two types of knowledge are argued to operate along a continuum on which declarative knowledge can transition into procedural, or automatised, knowledge over time with practice (Anderson, 1982, 1992, 2000; Anderson & Schunn, 2000; Carlson, 2003; DeKeyser, 1997; 1998, 2007b; R. Ellis, 1993; Sharwood Smith, 1981; Ullman, 2004, 2005). As mentioned previously, this continuum from declarative to procedural knowledge represents a break from the characteristics of implicit and explicit knowledge, which are believed to operate as two separate systems.

As part of the strong-interface position, the transition from declarative to procedural knowledge embodies a qualitative shift as a result of automatisation, also referred to as practice (DeKeyser, 2001). Automatisation refers to the process of "knowledge change from initial presentation of the rule in declarative format to the final stage of fully spontaneous, effortless, fast, and errorless use of that rule, often without being aware of it anymore" (DeKeyser, 2007b, p. 3). In this sense, the strong-interface position holds that automatised declarative knowledge, or procedural knowledge, can be seen as functionally equivalent to implicit knowledge. It should be reiterated, however, that automatised knowledge is not necessarily the same as implicit knowledge. Unlike implicit knowledge, a lack of awareness is "not a requirement for automaticity" (DeKeyser, 2007b, p. 4; see also R. Ellis, 1994; Ullman, 2005). In other words, it is possible to attain automaticity of a particular structure while being fully aware of the rules underlying it. As such, automaticity can characterise both implicit and explicit knowledge (DeKeyser, 1998, 2007b), with both types of learning involving a shift from more controlled to more automatic processing.

2.4.3. Weak-interface Positions

Other researchers (R. Ellis, 1993; 1994; N. Ellis, 2005) have proposed weak-interface positions for the interface between explicit and implicit knowledge. In R.

Ellis' (1993, 1994) weak-interface model, explicit knowledge can be converted into implicit knowledge under certain conditions. This weak-interface position distinguishes between developmental and variational features of language. Developmental features of language, such as negation or third person –s, are linguistic features whose acquisition appears constrained by universal orders of language development. In contrast, variational features, such as the copula "be", are linguistic features that do not appear constrained by acquisition orders. In R. Ellis' model,explicit knowledge can become implicit knowledge only with variational features of language and only when the learner is developmentally ready to incorporate this feature into his or her interlanguage system (R. Ellis, 2008). The learner's current existing knowledge thus serves "as a kind of filter that sifts explicit knowledge and lets through only that which the learner is ready to incorporate into the interlanguage system" (R. Ellis, 1994, p. 88). Drawing on Schmidt's noticing hypothesis (1990), this model also acknowledges that explicit knowledge and formal instruction can have a facilitating effect on implicit knowledge by drawing the learner's attention to relevant features in the input that would not have otherwise been noticed, as well as by "noticing the gap" (Schmidt & Frota, 1986); in other words, comparing what he or she has noticed with his or her own current interlanguage. In this view, explicit knowledge then facilitates the development of implicit knowledge by sensitising "the language processor so that it takes account of data in the input" (Schmidt & Frota, 1986, p. 98). However, the fact that noticing is considered a conscious process (Schmidt, 1990) suggests that this "taking account" by the language processor is a conscious, or explicit, process. In other words, what remains unclear within R. Ellis' model is how noticing, which is associated with conscious awareness, aids in the development of implicit, rather than explicit knowledge.

N. Ellis' weak-interface model (1993, 1994b, 2005, 2011, 2015; N. Ellis & Larsen-Freeman, 2006; see also the entries in N. Ellis, 1994a) shares features with both Krashen's (1982) non-interface position and R. Ellis' (1994) weak-interface position. Like Krashen, N. Ellis argues that implicit learning is primary[7]: Most learning is implicit, and most knowledge is tacit (2005). Also, as with the non-interface position, implicit and explicit knowledge represent distinct systems within N. Ellis' weak-interface model. Furthermore, explicit knowledge cannot be converted into implicit knowledge. This, however, should not be taken to mean that implicit and explicit knowledge operate independently of each other. According to N. Ellis, "conscious and unconscious processes are dynamically involved together in every cognitive task and in every learning episode" (2005, p. 340), including both input and output processing. As an illustration of the interplay between implicit and explicit knowledge during language use, N. Ellis argued that we rely primarily on automatic processing; however, we also draw upon explicit knowledge when automatic processes fail. Examples of this would include when we stumble while walking, or when communication breaks down. In this sense, explicit knowledge serves as a form of "collaborative conscious support" for more automatic processes (p. 308).

In addition to explicit knowledge ballasting implicit knowledge during language production, N. Ellis also maintained that there is a relationship between implicit and explicit knowledge as part of language development. A key feature of N.

---

[7] See also Reber (1990).

Ellis' (2005) weak-interface model is that it stresses the importance of attention in the development of both explicit and implicit knowledge. In this model, explicit knowledge develops as a result of conscious attempts at learning. Implicit knowledge, on the other hand, develops as a result of repeated exposure to exemplars within the input available to the learner. As a result of this repeated exposure, connections are strengthened within a connectionist framework via an unconscious tallying mechanism, resulting in the development of implicit knowledge (N. Ellis, 2005, 2011; Hulstijn, 2002). Of key importance here is that, although awareness is not necessary for tallying to take place within this model, attention is (N. Ellis, 2005). It is only after constructions have been detected, to borrow Tomlin and Villa's terminology, and then seeded within explicit memory that there is scope for implicit learning to take place. This differentiates N. Ellis' (2005) and R. Ellis' (1994) weak-interface models, in that explicit knowledge does not become implicit knowledge (as is the case in R. Ellis' model), although explicit knowledge does have a direct line of influence in the development of implicit knowledge.

2.4.4. Non-interface Revisited

Despite the arguments in favour of an interaction between implicit and explicit knowledge, not all researchers are convinced. Bialystok (1982, 1994) and Paradis (2004, 2009), for instance, have pointed out that evidence that might be interpreted as a shift from explicit to implicit knowledge can also be argued to indicate a shift in the control of processing. In other words, do changes in the degree of automaticity reflect the development of implicit knowledge, or simply the use of implicit knowledge? Paradis (1994, 2004, 2009) has further argued fervently that any changes in learners’ language production from explicit, declarative knowledge towards implicit, procedural competence should not be taken as evidence of an interface between explicit and implicit knowledge. For Paradis (2004), any shift in learners’ performance can be attributed to changes, or variation, in their reliance on one type of knowledge over the other. For example, in the early stages of language learning, learners may rely exclusively on explicit, metalinguistic knowledge. Over time, however, as their implicit knowledge develops, learners slowly come to rely less on explicit knowledge in favour of more automatic, procedural processes. As noted, Paradis (2004, 2009) holds that explicit and implicit knowledge develop independently throughout this process, and do not interact directly (2009).

This chapter has outlined the various theoretical models that have attempted to account for the development of implicit and explicit knowledge, as well as how these types of knowledge may interface in their development and use. The following chapter will shift the perspective towards research methodology, particularly with regard to how implicit and explicit knowledge have been operationalised and measured within cognitive psychology and SLA. Starting with the artificial grammar learning studies of Arthur Reber (1967), I will trace the use of artificial and semi-artificial languages, and outline how different tools for gauging awareness have been utilised in conjunction with these languages.

CHAPTER 3

METHODOLOGICAL BACKGROUND

Chapter 3 provides the methodological background to this thesis. This chapter first introduces the major paradigms on which the experiments in this thesis are based, namely the artificial grammar learning paradigm (Reber, 1967), the serial reaction time paradigm (Nissen & Bullemer, 1987) and the self-paced reading technique (Roberts & Felser, 2011). Following this, the chapter provides a critical overview of the types of stimulus arrays that are typically included in the experiments operating within the cognitive framework, specifically artificial languages, semi-artificial languages, and miniature language systems. At this point, the chapter shifts focus towards the measurement of implicit and explicit knowledge within the methodological frameworks listed above. Instruments such as verbal reports, subjective measures of awareness, and direct and indirect tests of learning are discussed in terms of their strengths and limitations in differentiating between implicit and explicit knowledge.

3.1. Artificial Grammar Learning

As in the field of SLA, implicit learning has been the focus of much research and debate in cognitive psychology, with interest in this area being revived after Arthur Reber's (1965, 1967) seminal investigation into the incidental learning of the statistical regularities of an artificial grammar system. In the training phase of this experiment, which was disguised as a study of rote memory, participants were presented with individual strings of letters that were between three and eight letters in length, such as *TPTXXVS*, on index cards. The training task required the participants to memorise these strings of letters to a criterion, which was measured by their ability to reproduce the strings from memory successfully by writing them down on a separate piece of paper. However, the participants were not informed that these strings followed a complex rule structure; in this case, they were generated according to an algorithmic model following a complex finite-state Markovian system (see Figure 3A below for a visual representation of how these strings were generated).

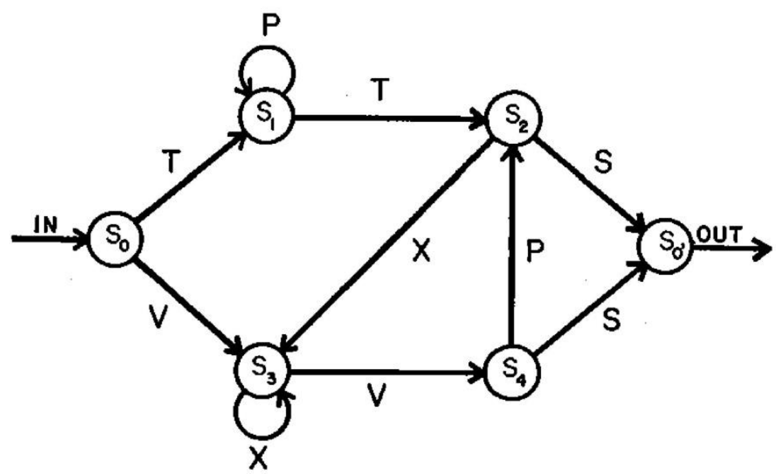


FIG. 1. Schematic state diagram of the grammar used to generate the grammatical stimulus items

Figure 3A. Reber's Markovian grammar (taken from Reber, 1967, p. 854).

Following the training phase, participants were given a surprise grammaticality judgement test (GJT). In this test, participants were presented with novel strings of letters, half of which were grammatical, in that they followed the same rule system as the strings of letters presented during the training phase, and the other half of which were ungrammatical, in that they violated the rule system in some way. For example, following the schematic representation of the grammar system in Figure 3A, the string *TPTXXVS* would be considered grammatical because it follows one of the possible paths through the model. On the other hand, the string *TTPSPXV* would be ungrammatical, because this particular construction cannot be formed by following the paths in the diagram.

The results of the testing phase indicated that participants were able to classify approximately 80% of the strings accurately. Reber argued that this significant performance was evidence that the act of memorisation in the training phase was sufficient for participants to develop an abstract mental representation of the underlying rule structure regulating the stimulus array. Furthermore, Reber argued that this process, for which he coined the term implicit learning, was an inductive and unconscious process that resulted in an unconscious knowledge base (see also Reber, 1993, 2011; Winter & Reber, 1994). However, a limitation of Reber's (1967) interpretation[8] of implicit learning is that his experimental design failed to include any measures to ensure that participants were not aware of the rule system. Thus, in the absence of data as to the conscious status of knowledge, Reber's claim that the knowledge was implicit appears to be an assumption, although it should be pointed out that later experiments (Reber & Allen, 1978) attempted to address this issue by including retrospective verbal reports as part of their design (see Section 3.5.2 below for a discussion of retrospective verbal reports).

Reber's (1965, 1967) study and subsequent experiments sparked considerable research in experimental psychology (for general overviews, see Cleeremans et al., 1998; Misyak, Goldstein, & Christiansen, 2012; Perruchet, 2008; Reber, 1993; Reber & Allen, 2009; Shanks, 2005), most of which followed what has become known as

[8] This criticism can be extended to many influential studies that have followed Reber's original methodology (see also Reber, 1976; Reber, Kassin, Lewis & Cantor, 1980; Reber & Lewis, 1977).

the artificial grammar learning (AGL) paradigm. In this paradigm, as in Reber’s study, participants are exposed to strings of letters, the construction of which is governed by a complex grammar system. Typically, participants are not informed about this underlying grammatical system, and instructions and tasks are designed to prevent participants from analysing the stimulus material (Perruchet, 2008; see Reber, Kassin, Lewis, & Cantor, 1980 for an exception). Following exposure, participants are told about the existence of the underlying system, and are then asked to distinguish sequences that have been generated by the same rule system from sequences that have not. Ordinarily, this involves the use of a grammaticality judgement test, although more recent studies have adopted alternate assessments (see the later discussion of serial reaction-time research). In studies that follow this paradigm, participants are generally able to perform at levels significantly above chance in the testing phase, even after a minimal amount of exposure to the training task (Reber, 1993). Awareness is typically assessed using retrospective verbal reports. Research into the learning of artificial grammars has been particularly robust, in that studies following the AGL paradigm (Allen & Reber, 1980; Brooks, 1978; Dienes & Altmann, 1997; Morgan & Newport, 1981; Perruchet & Pacteau, 1990; Reber, 1969, 1976) have been remarkably consistent in establishing a learning effect (Perruchet, 2008; Winter & Reber, 1994), to the degree that the AGL paradigm has been used as a demonstration of laboratory learning in psychology textbooks (Reber, 1993; Rebuschat, 2015b).

Although the fact that some learning typically takes place in AGL studies is well established, two of Reber's assertions, specifically that the participants in these experiments developed abstract rule knowledge and that this knowledge was unconscious in nature, have been fiercely contested in the literature (see Perruchet, 2008; Pothos, 2007 for reviews). Firstly, several studies in the early 1990s (Dienes, Broadbent, & Berry, 1991; Gómez & Schvaneveldt, 1994; Mathews, 1990; Perruchet & Pacteau, 1990; Perruchet, Gallego, & Savy, 1990; Perruchet, Gallego, & Pacteau, 1992; Servan-Schreiber & Anderson, 1990) challenged the claim that the training phase of Reber's (1967) early experiments resulted in abstract rule knowledge. These studies have demonstrated that participants become sensitive to the statistical regularities of bigrams and/or trigrams, that is to groups of two or three letters (such as *DV* or *XYV*), found within the longer strings of letters in the input (as in *DVPYXYV*). For example, Perruchet and Pacteau (1990) reported that the performance of learners who were trained only on bigrams was identical to those trained on entire strings. Similarly, data from Dienes et al. (1991) suggest that participants' classification of individual bigrams of longer strings fully accounted for their overall performance in judging complete strings.

In summary, these studies suggest that learners were relying on "piecemeal knowledge" of the strings of letters, rather than necessarily on an abstract knowledge of the linguistic system (Perruchet, 2008). This issue is partly connected with the use of a GJT as the chief measure of learning in these studies. As part of a GJT, participants are presented with individual strings of letters that they then have to indicate as being grammatical or ungrammatical. The issue here is that such a binary classification does not provide information concerning why participants judge some strings as grammatical and other strings as ungrammatical, nor can it identify which component, or sequence of components, participants utilise in the course of their judgements. To obtain this information, assessments would have to provide information about the participants' real-time processing of grammatical violations, information that a GJT is unable to provide.

Secondly, the degree to which the resulting knowledge in these studies is, in fact, implicit has been the focus of much debate in the cognitive sciences (Perruchet, 2008; Perruchet & Pacton, 2006; Shanks, 2005; Shanks & St. John, 1994). Within the AGL paradigm, the conscious status of knowledge has been measured via retrospective verbal reports (see Reber, 1993; Rebuschat, 2013, for overviews) in which participants are prompted at the end of the testing phase to verbalise any rules they noticed during the experiment. Typically, the results of AGL experiments indicate that participants are unable to verbalise the underlying rule structures, despite being able to classify novel strings of letters as grammatical or ungrammatical at levels significantly above chance level and/or above the performance of untrained control groups (Rebuschat, 2008). In other words, the knowledge reported by the participants is not sufficient to account for their performance (Dienes et al., 1991).

This significant performance in the absence of verbalisable rule knowledge has been argued as evidence of an unconscious, implicit, knowledge base (Reber, 1993). Critics of this interpretation have called into question the veridicality of retrospective verbal reports, specifically the sensitivity of this instrument in gauging awareness that occurred earlier in the experiment. In other words, the inability to verbalise the underlying rule system in retrospective verbal reports might simply reflect the limitations of this instrument, rather than tacit knowledge (Bialystok, 1979;

Brody, 1989; Dulany, Carlson, & Dewey, 1994; Leow, 2002, 2015a, 2015b; the limitations of retrospective verbal reports are discussed in more detail in Section 3.5.2). In light of these criticisms, some researchers (Dienes, 2004; Dienes & Scott, 2005; Sandberg, Timmermans, Overgaard, & Cleeremans, 2010) have turned to alternate instruments, such as subjective measures of awareness, to provide more valid data regarding the conscious status of knowledge. These measures of awareness are discussed further in Section 3.5.3 below.

3.2. Serial Reaction Time Methodology

To address some of the issues related to the use of GJT as the chief measure of learning, many studies in cognitive psychology have turned to online ("in real time") measures of learning. The most commonly utilised online methodology has been the serial reaction time (SRT) paradigm, or variants thereof (see Robertson, 2007, for a review). SRT tasks gained in popularity in the late 1980s as an alternative methodology to investigate the learning of sequential regularities (Cleeremans & McClelland, 1991; Lewicki, Hill, & Bizot, 1988; Nissen & Bullemer, 1987; Perruchet et al., 1990). Within this paradigm, participants track the location of stimuli presented on a monitor using response keys. For example, in an early study by Lewicki et al. (1988), the screen was divided into four quadrants by one vertical and one horizontal line. Participants were shown the letter "X," which appeared in one of the four quadrants. The participants were tasked to react to the appearance of the letter X by pressing the corresponding button on the keypad. The participants were not informed that the sequence of positions in which the letter appeared followed a regular pattern.

Thus, differences in reaction times to sequences that followed or violated this pattern serve as evidence of whether the sequence had been learned.

SRT research has proven to be a fruitful method for examining sequence learning and, when accompanied by measures of awareness, the role of awareness within this process (Williams & Rebuschat, 2013). However, the degree to which evidence from SRT tasks, which often involve examining reaction times to the location of a stimulus presented on a monitor, can be generalised to the learning of artificial, much less natural grammars, is questionable (Leow, 2015a, 2015b; Leow & Hama, 2013). Furthermore, from a methodological perspective, the traditional SRT paradigm does not appear to be immediately applicable to examining the learning of an artificial language. This is due to the fact that artificial languages are constructed according to constraints governing letter combinations and string length (Reber, 1967), and are not constructed with regard to the locations in which particular characters are presented on the monitor.

To overcome this methodological hurdle, some recent AGL studies have followed in the tradition of SRT research by adopting alternative methodologies that, as with SRT tasks, incorporate reaction times as a means of tracking learning. Both self-paced reading and/or eye-tracking tasks have been included among these methodologies (Amato & MacDonald, 2010; Karuza, Farmer, Fine, Smith, & Jaeger, 2014; Andringa & Curcic, 2015; Williams & Rebuschat, 2013). There are a number of methodological improvements in the use of these online tasks compared to traditional GJTs. First, as noted, online tasks can provide more fine-grained evidence about the aspects of the artificial grammar the participants have learned, and whether these are bigrams, trigrams or whole strings of letters by conducting fine-grained analyses of how reaction times, or eye movements, coincide with specific features of the input.

In addition, online measures have been argued to be more transfer-appropriate for the type of learning that occurs as a result of incidental exposure (Hulstijn, 2002, Whittlesea & Dorken, 1993). The concept of transfer appropriateness, which has roots in the depths of processing framework (Craik & Lockhart, 1972), holds that we are more likely to remember what we have learned if there is a match between the conditions under which the material is learned and the conditions under which it is tested. In other words, learning is more likely to transfer from training to testing when there is fidelity between the conditions of training and testing (see Hulstijn, 2003, 2013; Larson-Freeman, 2013; Lightbrown, 2008 for discussions of transfer-appropriate processing in SLA). In this regard, online tasks are advantageous in that the training and testing phases place identical, or very similar, processing demands on the learners. It has also been argued that online assessments are better indicators of implicit knowledge than are offline measures, because the constraints of processing in real-time mitigates the opportunities for participants to utilise explicit, metalinguistic knowledge in their judgements (Godfroid et al., 2015; R. Ellis, 2004, 2005; R. Ellis et al., 2009; Jegerski, 2014; Norris & Ortega, 2000; Perruchet, 2008). In summary, online measures of learning have been argued to be preferable to offline instruments because they reflect the learning that results from incidental exposure more reliably.

In the following section, I discuss the use of the self-paced reading task as a method for the collection of online data. I focus more on this type of task than on eye-tracking, as self-paced reading is a technique that is used in the research presented in this thesis.

3.3. Self-paced Reading Tasks

As noted above, one technique that allows for the collection of online data is the self-paced reading (SPR) task. Originally developed in the 1970s (Mitchell & Green, 1978), the SPR task is a computerised technique that has become increasingly popular in the field of psycholinguistics (see Jegerski, 2014; Keating & Jegerski, 2014; Marinis, 2003, 2010; Roberts, 2012a, for overviews). Although SPR tasks have primarily been applied in research examining various phenomena related to L2 processing, such as the processing of sentence ambiguities (Roberts & Felser, 2011), this technique has also been used to assess learners' sensitivity to grammatical violations during L2 processing, thus shedding light on learners' current levels of grammatical competence (Hopp, 2010; Jiang, 2004, 2007; Karuza et al., 2014; Sagarra & Herschensohn, 2011; Tokowicz & Warren, 2010; Williams & Rebuschat, 2013).

As part of the SPR task, which is also commonly referred to as a “moving windows” technique (Jegerski, 2014), participants read individual sentences in segments, either word-by-word or phrase-by-phrase. Typically, participants are given instructions to read the sentence as quickly as possible, but they are only presented with a single word or a single segment at a time. Participants must press a button, either on the keyboard or on the response pad, to progress to the next word or phrase in the sentence. The software package then records the time that the participant spends on each segment of the sentence. Learning is then operationalised according to the time it takes a participant to progress between the segments of the sentence. Sensitivity to grammatical violations in the form of increased reaction times is taken as evidence that the underlying grammatical rules have been acquired.

Although all SPR tasks tend to follow the basic model outlined, there are a few permutations of the basic SPR task in terms of the manner in which the stimulus sentences are displayed to the participants. Firstly, an SPR task can be cumulative; that is, once a word has been displayed, it remains visible to the participant until the end of the sentence or experimental trial (an illustration can be seen in Figure 3B below).

While	......	.....	......	......	......	.....	.....	.....	.....
While	the	.....	.....	.....	.....	.....	.....	.....	.....
While	the	band	.....	.....	.....	.....	.....	.....	.....
While	the	band	played	.....	.....	.....	.....	.....	.....
While	the	band	played	the	.....	.....	.....	.....	.....
While	the	band	played	the	beer	.....	.....	.....	.....
While	the	band	played	the	beer	pleased	.....	.....	.....
While	the	band	played	the	beer	pleased	all	.....	.....
While	the	band	played	the	beer	pleased	all	the	.....
While	the	band	played	the	beer	pleased	all	the	customers.

Figure 3B. Example of a cumulative self-paced reading design with word-by-word segmentation (Roberts & Felser, 2011).

Although this variant of the SPR task might appear to be reflective of normal, incremental reading processes, cumulative SPR tasks are problematic for experimental purposes because participants tend to develop reading strategies in which they reveal several segments of the sentence at a time before reading them (Ferreira & Henderson, 1990; Jegerski, 2014; Just, Carpenter, & Wooley, 1982; Marinis, 2003, 2010). Thus, the construct validity of the experimental design is threatened, since the time measurements recorded for each segment might not accurately reflect the time the participant spent reading that segment in particular.

An alternative to a cumulative display is a non-cumulative presentation. Within such a presentation, only one segment is visible at a time and, when a new segment is revealed, the previous segment is hidden (see Figure 3C below).

.....	.....	While	.....	.....
.....	.....	the	.....	.....
.....	.....	band	.....	.....
.....	.....	played	.....	.....
.....	.....	the	.....	.....
.....	.....	beer	.....	.....
.....	.....	pleased	.....	.....
.....	.....	all	.....	.....
.....	.....	the	.....	.....
.....	.....	customers	.....	.....

Figure 3C. Example of a centred non-cumulative reading design with word-by-word segmentation (Roberts & Felser, 2011).

Non-cumulative presentations have included both centred presentations in which each word is presented in succession in the same location at the centre of the screen, as well as linear presentations in which the sentence is presented from left to right (or right to left, depending on the nature of the language being investigated) on the screen without any overlap in presentation (Figure 3D below). As others (Jegerski, 2014) have noted, the ecological validity of the centred version of the non-cumulative SPR task has been challenged on the grounds that the manner of presentation might not reflect normal reading processes, such as the “left to right” reading processes of participants with an English L1. Consequently, linear presentations are preferred because they are closer to what might be considered a “normal” reading than is a centred presentation. Given the shortcomings of the cumulative and centred non-cumulative versions of the SPR task, nearly all recent SPR studies have opted for a linear non-cumulative version of the SPR task in their experimental design (Jegerski, 2014).

While	......	.....	......	......	......	.....	.....	.....	.....
.....	the	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	band	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	played	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	the	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	beer	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	pleased	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	all	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	the	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	customers.

Figure 3D. Example of a non-cumulative self-paced reading design with word-by-word segmentation (Roberts & Felser, 2011).

Two recent studies operating in the AGL tradition (Amato & MacDonald, 2010; Karuza et al., 2014) have incorporated SPR tasks in their examinations of the statistical learning of artificial grammars. First, Karuza et al. (2014) examined the learning of an artificial grammar of shape sequences. In this study, the stimulus material consisted of glyphs, namely symbols from Ge'ez script (the writing system found in Ethiopia and Eritrea). These glyphs were chosen because they are visually distinct and represent a writing system with which participants would be unfamiliar. Similarly to the letter strings utilised in early AGL research (Reber, 1967), the order of these symbols was governed by a complex rule system. The exposure phase consisted of 432 trials in which triplets of the glyphs were presented via a self-paced exposure phase. These triplets contained both "predictable" combinations, such as combinations of glyphs that co-occurred at high rates of probability across the training phase, and "unpredictable" elements, or glyphs that co-occurred at low rates of probability across the training phase. Built into the exposure phase were 144 intermittent "catch" trials, in which participants were asked if the previous set of triplets contained a particular symbol. The purpose of the catch trials was to ensure that participants were paying attention throughout the training phase and were not simply clicking through the experiment. Following exposure, the participants completed a familiarity judgement task in which they were presented with triplets in their entirety. Participants then had to indicate whether the triplet seemed familiar; in other words, whether it had appeared in the training phase. Data collected over the course of the experiment indicated that participants' reaction times decreased quickly for triplets containing predictive glyphs, but reaction times did not decrease for non-predictive sequences. In other words, participants responded more quickly on the SPR task for predictive sequences than for non-predictive sequences. These results suggest that the participants rapidly derived some knowledge of the underlying patterns of the artificial system based on the statistical probabilities that certain combinations of glyphs occurred in the input.

In another study, Amato and MacDonald (2010) measured the learning of a miniature artificial language via two separate testing instruments: a self-paced reading task intended to measure implicit knowledge, and a sentence completion task argued to reflect explicit knowledge. The artificial language used in this experiment was "miniature" in that it only consisted of 18 words: three animate nouns designated as the subject of the sentences, three inanimate nouns designed as the object of sentences, three verbs, five adjectives of colour, three adjectives describing patterns or markings (such as stripes and spots), and one preposition. Sentences like *Veek pim mog minada sarp skod* that corresponded to events depicted in cartoon images were created. Each cartoon image contained one cartoon monster (always the subject) acting on one inanimate, nonsense object. The frequency with which the individual combinations of monsters, verbs, and objects appeared was experimentally manipulated so that certain combinations occurred more frequently than did others, with certain combinations occurring with a high rate of frequency and others with an extremely low rate.

The training phase in this study consisted of four blocks. The first three consisted of vocabulary training/sentence learning in which participants had to match the cartoon images with one of two sentences displayed on the screen. The fourth block in this study consisted of a non-cumulative, centred, self-paced reading task. Following each sentence in the SPR task, participants were presented with a picture and were asked to indicate whether the picture matched the previous sentence. Following 36 training trials, participants completed 36 test trials in which their reaction times to rare versus frequent combinations of the artificial language were compared. Although no measurement of awareness was included in this study, this task was argued to assess more automatic, implicit knowledge. Following the self-paced reading task, participants completed two further tasks designed to access any explicit knowledge acquired during the training phase of the experiment. The first was a sentence completion task in which a sentence with a missing word, such as "*Veek pim mog minada* _____ *skod*", was presented on the monitor. Participants had to choose between three options, "*gorm*, *sarp*, or *clate*" to complete the sentence. The second task presented participants with a picture in which one of the objects had been removed. Participants were asked to select one of three options to indicate which object had been removed.

The results of the explicit tasks of this experiment revealed no significant learning effect across either the sentence-completion or picture-completion tasks. The results of the self-paced reading task did, however, find significant differences in the participants' reaction times between frequent and rare combinations. The results of the SPR task can thus be interpreted as suggesting that the participants had developed some sensitivity to the underlying regularities of the artificial language system. Furthermore, the results from this study support arguments that behavioural measures of learning, such as reaction time tasks, are a more sensitive measure of learning than are traditional learning tasks, such as the sentence completion task used in Amato and MacDonald (2010; see also Hulstijn, 2002 for a discussion).

In summary, the artificial grammar learning paradigm, as established by the work of Reber (1967) has generated considerable interest, and controversy, in the cognitive sciences. This ongoing debate has largely focused on the degree to which statistical regularities can be inducted as a result of exposure to input, as well as whether the resulting knowledge is conscious or unconscious in nature. Furthermore, from a methodological perspective, the AGL paradigm has proven to be a robust methodological framework for experiments examining incidental learning within the cognitive sciences. However, for researchers operating in other domains, such as second language acquisition, one important question is the degree to which the findings from AGL research, which has investigated the learning of meaningless strings of letters, can be generalised to more complex stimulus domains, such as natural languages (Robinson, 2010; Schmidt, 1994a). The following section will review literature that is relevant to this issue—specifically, the discussion surrounding the limitations of artificial grammars in the study of SLA, and attempts by researchers to address these limitations.

3.4. Of Artificial Grammars and SLA

In an early review, McLaughlin (1980) pointed out that artificial language research has focused on two broad areas of inquiry. The first is concerned with how learning occurs and the general learning mechanisms underlying this process. The second hinges on the intrapersonal and situational variables that mediate the process of learning. Given the focus and scope of this research agenda, it might seem that the results from artificial language studies should be relevant to all fields related to education and learning, including SLA. However, the validity and generalisability of the results of artificial language experiments have been challenged in SLA, particularly due to the differences in complexity between artificial language and natural language systems (DeKeyser, 1994; N. Ellis, 1999; McLaughlin, 1980; Schmidt, 1994b; Winter & Reber, 1994).

These differences in complexity have not resulted in the abandonment of artificial language research within SLA, and recent years have witnessed an increasing interest in this methodology, presumably because of the significant advantages offered by using artificial systems. Firstly, potentially confounding variables, such as prior knowledge, can effectively be accounted for, and input variables such as frequency of the target structure can be readily manipulated. In short, artificial systems provide the experimenter with complete control over the characteristics of the input (Cook, 1988). Secondly, to return to a previous point, the artificial language paradigm has generated very robust and easily replicable findings since the 1960s (Braine, 1963, 1966; Moeser & Bregman, 1972; Reber, 1967; Segal & Halwes, 1965, 1966; Smith, 1966). Furthermore, the general methodology has proven to be flexible and adaptable to different experimental contexts (see Gómez & Gerken, 2000 and Ziori & Pothos, 2015, for reviews of artificial language and artificial grammar research, respectively).

Given the methodological advantages of the AGL paradigm, a number of SLA studies have taken early criticisms of artificial languages (Schmidt, 1994a) into account and have utilised artificial languages with natural referents in the form of miniature language systems or semi-artificial languages in an attempt to bridge the gap between construct validity and experimental control. These studies have contributed to the body of knowledge on a wide range of issues related to psycholinguistics and SLA, including the degree to which various aspects of L2 grammar can be acquired under incidental learning conditions (Alanen, 1995; DeKeyser, 1994, 1995; Hama & Leow, 2010; Leung & Williams, 2011, 2012; Rebuschat, 2008; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, 2013, 2015;

Rebuschat & Williams, 2006, 2009, 2012; Robinson, 2002, 2005, 2010; Williams, 2004, 2005; Williams & Kuribara, 2008).

3.4.1. Miniature Language Systems

Miniature language systems, simply put, represent simplified versions of a single natural language. Typically, these include a restricted number of lexical items, as well as controlled morphosyntactic features. For example, a study by Robinson (2002) utilised a miniature Samoan system, consisting of one article, 11 verbs, and 15 nouns, to examine the acquisition of L2 morphology, specifically locative, ergative and noun incorporation case markers. Other studies have successfully developed miniature systems across a wide range of artificial languages (Andringa & Curcic, 2014; de Graaf, 1997; DeKeyser, 1995, 1997; MacWhinney, 1983) as well as natural languages, including Finnish (Alanen, 1995), Russian (Brooks & Kempe, 2013), Samoan (Robinson, 2002, 2005, 2010), and Basque (Tagarelli, 2014; Tagarelli et al., 2014).

In an exemplary series of experiments, Patricia Brooks, Vera Kempe, and their associates (Brooks & Kempe, 2013; Brooks, Kempe, & Donachie, 2011; Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2001; Kempe, Brooks, & Kharkhurin, 2011; Kempe, Brooks, & Pirott, 2001) have used several variations of a miniature linguistic system based on Russian to investigate issues related to the early stages of both first and second language acquisition. For example, Kempe and Brooks (2001) tested the degree to which diminutive endings might facilitate the acquisition of Russian gender agreement patterns for adult L2 learners. Drawing inspiration from L1 acquisition research that has revealed that child-directed speech (CDS) might facilitate the acquisition of L1 phonology, Kempe and Brooks (2001) set out to explore this effect among adult L2 learners (see also Kempe, Brooks, & Gillis, 2007 for a review of research on the benefits of diminutives for language acquisition). Building on their previous research on L1 acquisition (Kempe et al., 2011; see also MacWhinney, 1995), this study investigated whether characteristics of CDS in Russian might facilitate the acquisition of aspects of L2 Russian grammar.

As Kempe and Brooks (2001) noted, Russian diminutive forms are appealing in CDS because they carry connotations of endearment, affection, and smallness. These diminutives, however, also increase morphophonological transparency in that they help to regularise suffixes related to Russian gender marking. For example, in Russian, feminine nouns tend to end in *-a*, and masculine nouns tend to end in a non-palatised consonant in their nominative, simplex forms. Examples of these are lisa [fox] and ptica [bird] for feminine nouns and zhuk [beetle] and slon [elephant] for masculine nouns. These four nouns can be said to be transparent with regard to gender because their endings in the simplex forms make their gender category readily available. Despite having some gender marking regularities, several masculine and feminine nouns are non-transparent in their nominative, simplex forms. When confronted with these nouns in their simplex forms, it would be unclear to a learner whether they should be classified as masculine or feminine. Examples of this are korabl' [boat], which is masculine, and krovat' [bed], which is feminine. However, diminutive forms in Russian are highly regular in their endings. Feminine nouns in their diminutive form typically end in *-ka*, while masculine nouns in their diminutive form tend to have endings such as *-ik*, *-ek*, or *-ok* (Kempe & Brooks, 2001). To return to these examples, lisa [fox], ptica [bird], and krovat' [bed] become lisichka, ptichka, and korvatka, respectively. The masculine nouns zhuk, slon, and korabl become zhukok, slonik, and korablik, respectively. As can be seen, the diminutive forms of Russian nouns are longer than their simplex forms; however, the diminutive endings regularise the noun endings within genders, so that that all diminutive forms are transparent in their gender marking, even for nouns that are non-transparent in their simplex forms.

As previously stated, the purpose of Kempe and Brook's (2001) study was to investigate whether characteristics of the Russian CDS, in this case diminutive noun forms, can facilitate the acquisition of Russian grammar, in this case gender agreement, among adult L2 learners of Russian. For this study, they utilised a miniature linguistic system, based on Russian. This system consisted of 40 Russian nouns, 20 masculine nouns and 20 feminine nouns. Fifteen of the masculine nouns and 15 of the feminine nouns were transparent with regard to gender in their nominative, simplex forms. Five of the feminine nouns and five of the masculine nouns were non-transparent. All the nouns in this study were also matched with a diminutive form. There were two groups in this study, both of which received training on the miniature system over four training sessions spread over a period of 10 days. The difference between the two groups was that one was exposed to the simplex forms of the nouns only, whereas the other received training only on the diminutive forms.

The results of this study indicate that the simplex group outperformed the diminutive group in terms of learning the correct pronunciation of Russian words.

This result is perhaps unsurprising, given that the simplex forms of Russian words are much shorter and would thus be easier to learn to pronounce correctly. Nonetheless, the diminutive group significantly outperformed the simplex group in learning the gender categorisations that occurred in the training phase and in applying this knowledge to new nouns in the testing phase. This result suggests that exposure to the morphological regularities of the diminutive forms in the training phase helped the participants in the diminutive group to learn the genders of the nouns, and to then generalise this knowledge. These results are in line with other research (Monaghan, Christiansen, & Fitneva, 2011) that has indicated that greater systematicity among word forms facilitates the ability to group words according to category (such as gender), whereas greater arbitrariness between word forms of the same category promotes learning the meaning of individual words.

In summary, Kempe and Brooks' (2001) use of a miniature Russian system allowed an investigation into the early stages of language acquisition with a high degree of experimental control. By limiting the number of nouns in the study, they were able to have complete control over variables that could potentially impact on learning, such as the frequency of exposure to individual lexical items. In addition, limiting the scope of the stimulus array allowed for more practical experimentation because it was then possible for participants to learn the target structures within a shorter time than would be expected when exposed to an unsimplified natural language system.

3.4.2. Semi-artificial Language Systems

As an alternative to miniature systems, such as the miniature Russian system in Kempe and Brooks' (2001) study, semi-artificial systems have also proven useful in investigating the initial stages of language acquisition. Semi-artificial languages combine aspects of two or more natural languages. Typically, the lexis in semi-artificial languages is presented in the L1, or a well-known L2, of the participants of the study. This ensures that participants can pay attention to the meaning of the stimuli, and it greatly reduces the need for vocabulary pre-training or multiple training sessions, as was the case with Kempe and Brooks (2001) above. While the vocabulary of semi-artificial languages is presented in a familiar language, this lexis is also often combined with the morphosyntactic features of a different language, which represent the learning target of the experiment. For example, in an experiment with native speakers of Dutch, Hulstijn (1989) used a semi-artificial language consisting of Dutch vocabulary with the addition of artificial morphemes and an artificial word order. Rebuschat (2008; see also Rebuschat & Williams, 2006, 2009, 2012) developed a language system consisting of English words with German syntax to investigate the incidental learning of L2 word order.

Another example of a study using a semi-artificial language is Williams and Kuribara’s (2008) investigation of the acquisition of L2 Japanese word order. For this study, Williams and Kuribara developed a semi-artificial language of English words and Japanese syntax, with some of the English words also being marked by a Japanese case marker according to their function within the sentence; for example, *Student-ga dog-ni what-o offered?, Vet-ga injection-o gave,* and *That sandwich-o*

John-ga ate. As can be deduced from these examples, the *-ga* marker indicates the subject of the sentence, *-o* the object, and *-ni* the indirect object. In the experimental phase, participants were exposed to target sentences both visually and auditorily via a plausibility judgement task. Following the common procedure within the incidental paradigm, the participants were not informed that they would later be given a test on the underlying syntactic system. The learning of the underlying syntactic rule system was assessed using a surprise GJT. The results demonstrated a positive learning effect for the target syntactical structures, suggesting that incidental learning conditions can lead to the acquisition of L2 word order.

In summary, the use of miniature systems and semi-artificial languages has grown out of the cognitive tradition, which is willing to sacrifice some degree of external validity in exchange for reliability and experimental control (Hulstijn, 2015; Hulstijn et al., 2014). Artificial systems allow researchers to focus their experiments on the acquisition of specific linguistic features (such as gender categories in Kempe & Brooks, 2001; word order regularities in Williams & Kuribara, 2008), and effectively control for participants' prior knowledge. As such, miniature systems and semi-artificial languages appear to be particularly well-suited to investigating the initial stages of language learning, most notably the effects of first-exposure on the acquisition of various aspects of L2 grammar (Grey et al., 2014, 2015).

Looking critically at miniature systems and semi-artificial languages, miniature systems could be argued to be the more valid alternative because they more closely reflect the complexity of a natural language system. From a methodological point of view, however, miniature systems come with several drawbacks. In particular, as the lexis in miniature systems is in an entirely unfamiliar L2, these studies require either an extensive vocabulary pre-training phase (Robinson, 2002) or a more prolonged exposure phase (Kempe & Brooks, 2001) to allow participants enough time to learn the vocabulary used in the experiment. By contrast, semi-artificial languages, as noted above, typically utilise lexis from the participants L1, thus obviating the need for any vocabulary pre-training. In this regard, semi-artificial languages have a clear methodological advantage over miniature systems because they provide more ease of experimentation, and allow for the possibility of an entire experiment being completed in a short, single session. Semi-artificial languages appear particularly well suited to the AGL methodology and, when taken together, can provide insight into the earliest stages of language learning.

This section has provided a brief overview of two common types of stimulus arrays typically used in studies investigating the construct of awareness in SLA. The next section continues to focus on research methodology and outlines several instruments that are used to gauge awareness.

3.5. Measuring Implicit and Explicit Knowledge

3.5.1. Online versus Offline Measures of Awareness

Several instruments have been developed to tap into awareness, both in the cognitive sciences and in SLA. These instruments can be roughly categorised according to the point at which they are administered within the experimental design. Online, or concurrent, measures of awareness are administered concurrently; that is, during the exposure phase of the experiment (Bowles, 2010a, 2010b; Bowles & Leow, 2005; Leow, 2000, 2015a, 2015b). Think-aloud protocols (Bowles, 2008;

Hama & Leow, 2010; Leow, 1998, 2000; Rebuschat et al., 2013, 2015) and, more recently, eye-tracking software (Godfroid et al., 2015; Godfroid & Schmidtke, 2013; Godfroid & Winke, 2015) are examples of online measures of attention and awareness that have been used within the field of SLA. In contrast to online measures, offline, or retrospective, measures of awareness are administered following the exposure phase, typically following the testing phase at the conclusion of an experiment (Payne, 1994; Rebuschat, 2013). Of the offline instruments that have been used to measure awareness retrospectively, verbal reports have been the instrument most commonly used in both the cognitive sciences (Seth et al., 2005) and SLA (Rebuschat, 2013). In addition to verbal reports, recent years have also seen the use of subjective measures of awareness (Dienes & Scott, 2005) as alternate offline instruments.

It is important to note that both concurrent and retrospective measures seek to provide information about the participants' level of awareness during the experiment. However, the fact that these instruments are administered at different points in the learning process suggests that they provide data on different constructs (Hama & Leow, 2010; Leow, 2015a, 2015b). As noted, concurrent measures of awareness are taken as incoming information is being encoded (Hama & Leow, 2010). As such, the data provided by these measures can only be used as evidence of a participant's level of awareness at the same time the measurement was taken—in other words, during the encoding process. Thus, concurrent measures are best suited to experiments investigating issues related to implicit versus explicit learning, such as the role of attention and awareness during the learning process (see the work of Ron Leow, 1997, 1998, 2000, for examples). Conversely, the degree to which retrospective measures, which are taken at the end of the experiment, can provide accurate insight into the role of attention and awareness during the encoding process is questionable given that much of the information that is noticed during the encoding stage is unlikely to be remembered at a later stage in the experiment (Leow, 2002; Leow & Hama, 2013; Hama & Leow, 2010; Schmidt, 1994b; Shanks & St. John, 1994). Retrospective measures can, however, provide some information of the participants' level of awareness during the retrieval of stored knowledge (Leow, 2015b). In other words, retrospective measures are best suited to experiments investigating issues related to the product of learning, such as implicit versus explicit knowledge. As the focus of this thesis is on the development of implicit and explicit knowledge, only instruments related to measurement of the product of learning are discussed. For reviews of concurrent measures of awareness, see Bowles (2010b), Godfroid and Winke (2015), and Bowles and Leow (2005).

3.5.2. Retrospective Verbal Reports

Historically, retrospective verbal reports have been the most common instrument for assessing awareness in empirical research studies within the cognitive sciences (Seth et al., 2005; Leow, 2015b; Reber, 1993; Rebuschat, 2013). As part of the verbal reports, participants are asked at the end of the experiment to verbalise any rules or regularities they might have noticed during the experiment. Knowledge is typically considered implicit if participants are able to demonstrate knowledge (such as via their performance on a grammaticality judgement test), but cannot verbalise the underlying rule guiding their performance (Rebuschat & Williams, 2012; Robinson, 1997, 2002, 2005, 2010; see also Rebuschat, 2013 for a review). The results of verbal reports can be used to group participants dichotomously as being "aware" or "unaware" of the underlying rule system (Grey et al., 2014; Williams, 2005) or along a continuum based on the amount of information they are able to report (Brooks & Kempe, 2013).

One study that used retrospective verbal reports was Williams' (2005) investigation into the incidental learning of a semi-artificial noun-determiner system. In this study, participants were exposed to a relatively simple system that consisted of four determiners (*gi*, *ro*, *ul*, and *ne*) that were paired with a variety of English nouns (for example *ro* bull). Participants were told that the determiners functioned in the same manner as English determiners, except that these determiners also carried meaning in relationship to distance, with *gi* and *ro* being used for near objects and *ul* and *ne* being used for objects that were far away. What the participants were not told was that the determiners also encoded animacy, with *gi* and *ul* being used with animate objects and *ro* and *ne* with inanimate objects.

In the training phase of the experiment, participants listened to a sentence containing the noun-determiner system (*As I was passing I knocked over ro vase*). They were then asked to judge whether the determiner carried meaning for near or far, to repeat the sentence aloud in its entirety, and then to form a mental image of the situation described in the sentence. The first section of the testing phase of the experiment consisted of a sentence completion task. In this task, participants read the first half of a sentence, such as *the drunk fell over*, and then had to choose which of two options best completed the sentence; for example, *ro stools* or *gi stools*.

Following this task, the participants were interviewed to establish who had become aware of the underlying animacy rule system. The unaware participants were then given the same test sentences again, this time with the instruction to seek out the rules. Participants were then interviewed again to determine if they had become aware of the underlying system. The results of this experiment indicated that participants who remained unaware of the underlying rule system were still able to perform at levels significantly above chance on the sentence completion task. Williams has argued that these results provide some evidence that learning without awareness is possible.

Williams' (2005) study sparked considerable debate within SLA (Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leow & Hama, 2013; Leung & Williams, 2011, 2012, 2014; Rebuschat et al., 2013, 2015). A point of particular controversy was Williams' claim that the findings of his 2005 study pointed towards a dissociation between learning and awareness. The issue here was that Williams' (2005) study relied on retrospective verbal reports as the sole measure of awareness. As noted previously, the degree to which retrospective verbal reports can accurately plumb whether learning has occurred implicitly has been hotly contested (Hama & Leow, 2010; Leow, 2015a, 2015b; Leow & Hama, 2013). This is not to say, however, that Williams' (2005) study would have necessarily escaped criticism if he had simply asserted that the knowledge resulting from the training phase was implicit, rather than that learning had taken place implicitly. Several researchers (Leow, 2002; Reingold & Merikle, 1988; Schmidt, 1994a; Shanks & St. John, 1994) have challenged the interpretation of a lack of verbalisation in retrospective verbal reports as indicative of implicit knowledge. For instance, participants may choose to withhold information for a number of different reasons, such as lacking the meta-language needed to describe linguistic rules or regularities (Schmidt, 1994a). As such, interpreting a lack of verbalisation as being indicative of implicit knowledge should be viewed with some caution (see Leow, 2015b for a review).

In an attempt to address this lack of sensitivity, some researchers, such as Williams (2005), have modified the general procedure of retrospective verbal reports to allow participants multiple opportunities to uncover and report the underlying rule system. Other studies (Leung & Williams, 2011, 2012, 2014; Rebuschat & Williams, 2012) have encouraged participants to guess in the event that they do not know the underlying rule system. Despite such attempts to address the shortcomings of retrospective verbal reports, relying on these instruments as the sole measure of awareness remains problematic. As noted, the absence of verbalisation in itself is not compelling evidence that knowledge is implicit. Conversely, the presence of verbalisation might not necessarily indicate that the knowledge is completely explicit (Rebuschat et al., 2015).

3.5.3. Subjective Measures of Awareness

More recent experiments have begun to triangulate data by employing subjective measures of awareness in addition to retrospective verbal reports (see Dienes, 2008b, 2010, 2013; Rebuschat, 2008, 2013 for overviews). Subjective measures of awareness are predicated on the results of subliminal perception research (see Rebuschat, 2013; Timmermans & Cleeremans, 2015 for summaries), in which it has been suggested that the conscious/unconscious distinction should be defined in terms of subjective thresholds as opposed to objective thresholds (Cheesman & Merikle, 1986; Merikle, 1982, 1992; Reingold & Merikle, 1993; Smallwood & Schooler, 2009; Velmans, 1999, 2009a, 2009b). Simply defined, objective methods of assessing the conscious status of knowledge rely on carefully constructed tasks that are believed to reflect conscious or unconscious knowledge. For example, within SLA, it has been argued that learners are more likely to draw on implicit knowledge when faced with elicited imitation tasks and timed grammaticality judgement tests, as opposed to untimed grammaticality judgement tests that educe more explicit knowledge (R. Ellis et al., 2009; Godfroid et al., 2015).

Unlike objective methods, which determine whether knowledge is implicit or explicit based on task performance alone, subjective methods of awareness compare performance on a task with the degree to which the participants are able to verbalise the knowledge guiding their performance. Subjective methods can be further classified according to whether they assess first-order levels of awareness or metacognitive levels of awareness (Timmermans & Cleeremans, 2015). First-order awareness represents knowledge that the learner is able to report. An example is the ability to state the underlying grammatical rule in a retrospective verbal report. By contrast, metacognitive levels of subjective awareness refer to awareness of the existence of knowledge—for example, the learner is aware that he or she has learned something, but is unable to verbalise exactly what it is he or she has learned. An example of subjective metacognitive awareness in practice might occur when a learner is absolutely certain that an item on a GJT is incorrect, despite not knowing why, and thus being unable to verbalise the knowledge guiding his or her judgement.

Building further on the distinction between first-order and metacognitive subjective awareness, Dienes and colleagues (Dienes, 2004, 2008b, 2010, 2013; Dienes, Altman, Kwan, & Goode, 1995; Dienes & Berry, 1997; Dienes & Scott, 2005) have attempted to demarcate the threshold between implicit and explicit knowledge in terms of metacognitive subjective experience within implicit learning research. Within this body of research, it has been argued that the results of subjective measures of awareness reflect two different types of knowledge: structural knowledge and judgement knowledge (see Dienes, 2004, 2008b, 2010, 2013; Rebuschat, 2013; Rebuschat et al., 2013 for discussions). According to Dienes (2004), structural knowledge represents knowledge that participants acquire about the underlying structure of the stimuli during the experiment; for example, knowledge of the sequence of letters in an AGL experiment. Later, when confronted with a new string (in the testing phase of an experiment), participants are able to develop a new type of knowledge, judgement knowledge, whereby they judge whether the structure of the new item shares the same structural characteristics as the items in the training phase (Dienes, 2013). Although it is assumed that conscious structural knowledge leads to conscious judgement knowledge (Dienes & Scott, 2005), it has been argued that unconscious structural knowledge can lead to judgement knowledge that is either conscious or unconscious (see Rebuschat et al., 2013 for a discussion). If we take the example of performance on a grammar test, unconscious structural knowledge and unconscious judgement knowledge would be realised in situations in which someone truly believed him or herself to be guessing, yet his or her performance indicated that he or she possessed knowledge of the grammatical items being tested. On the other hand, unconscious structural knowledge and conscious judgement knowledge would manifest as "fringe feelings" of correctness (Norman, Price, Duff, & Mentzoni, 2007); in other words, the person would have a feeling (intuition) that a particular item is grammatical or ungrammatical, but would not know the reason (Dienes, 2010). Thus, it is possible for subjective measures of awareness to indicate that knowledge is fully conscious (conscious structural knowledge and conscious judgement knowledge), fully unconscious (unconscious structural knowledge and unconscious judgement knowledge), or partially unconscious (unconscious structural knowledge and conscious judgement knowledge). In other words, subjective measures of awareness allow for a finer-grained interpretation of the conscious status of knowledge than can be inferred from retrospective verbal reports.

There are two main types of subjective measures: source attributions and confidence ratings. Source attributions, which are designed to assess judgement knowledge, can be obtained alone (Dienes, 2004; Marsden, Williams, & Liu, 2013) or in addition (Dienes & Scott, 2005; Rebuschat & Williams, 2012) to confidence ratings. Source attributions ask participants to identify the source of their judgement by identifying the basis of their decision (for example, guessing, intuition, memory, or rule knowledge). Knowledge can be considered implicit if the performance is significantly above chance levels when participants are basing their decisions on intuition or when they believe themselves to be guessing (Dienes et al., 1995). Above-chance performance when participants believe themselves to be guessing is called the guessing criterion. Confidence ratings, which can be used to assess the conscious status of judgement knowledge (Dienes et al., 1995), ask participants to indicate their level of confidence (no confidence, somewhat confident, very confident, absolutely certain) for each decision they make during the testing phase (judging a sentence to be grammatical or ungrammatical). Dienes et al. (1995; see also Sandberg et al., 2010; Dienes, 2008; Rebuschat, 2013; Overgaard, Timmermans, Sandberg, & Cleeremans, 2010, for discussions) hold that knowledge can be considered unconscious if there is no statistically significant relationship between the level of confidence and performance. This is called the zero-correlation criterion.

Several recent studies within SLA have included subjective measures of awareness in their experimental design (Grey et al., 2014; Hamrick, 2013; Hamrick & Rebuschat, 2012, 2013; Rebuschat, 2008; Rebuschat & Williams, 2006, 2009, 2012; Serafini, 2013). For example, a recent study by Rebuschat and Williams (2012) investigated the acquisition of L2 German word order under incidental learning conditions. In this study, native English participants were exposed to a semi-artificial language comprised of English lexis, but governed by a syntactical rule system based on German word order; an example is *Usually defended Brian many shots during his matches.* The exposure phase required participants to listen to a sentence, repeat the sentence aloud, and to then make a semantic plausibility judgement based on the content of the sentence. Following the exposure phase, participants were given a surprise grammaticality judgement test in which they were asked to classify new sentences as grammatical or ungrammatical. Alongside each judgement on the GJT, participants also completed confidence intervals and source attributions. In addition, retrospective verbal reports were administered at the end of the testing phase. The results demonstrated a clear learning effect, indicating that L2 word order can be acquired through incidental exposure. Furthermore, following analyses of the subjective measures of awareness and retrospective verbal reports, the authors argued that this learning effect was partly driven by unconscious knowledge of the underlying syntactic rule system.

In summary, Rebuschat and Williams (2012) provided some evidence that L2 word order can be acquired as a result of incidental exposure, and that this incidental exposure leads to knowledge that is at least partly implicit in nature. This finding for partial implicit knowledge in Rebuschat and Williams (2012), and other studies that have utilised subjective measures of awareness (Hamrick, 2013; Hamrick & Rebuschat, 2013; Marsden et al., 2013; Serafini, 2013; Tagarelli, Borges Mota, & Rebuschat, 2011, 2015), illustrates the advantage of using subjective measures of awareness in conjunction with retrospective verbal reports. As noted, subjective measures of awareness are argued to elicit both first-order and metacognitive awareness, whereas retrospective verbal reports can reveal only first-order awareness. Thus, subjective measures of awareness appear more suited to revealing low levels of awareness that might manifest as a relationship between the level of confidence and accuracy during performance. Also, retrospective verbal reports are conducted at the end of the experiment, typically at the end of the testing phase. Subjective measures, on the other hand, are taken on a question-by-question basis throughout the training phase, thus allowing a finer-grained look at how awareness is guiding performance, as well as whether participants develop awareness over the course of the testing phase of the experiment (Hama & Leow, 2010; Rebuschat, 2008).

Despite the benefits of using subjective measures of awareness, reservations have been expressed regarding the interpretation of source attribution data. In particular, the interpretation of "intuition" to reflect native-speaker-like implicit knowledge (Dienes, 2008b; Rebuschat & Williams, 2012) has been challenged on the grounds that intuition, at least the phenomenology of the intuition possessed by native speakers of a language, develops slowly over a prolonged period (DeKeyser, 1997). It is thus questionable whether second language learners can develop such intuition following a minimal amount of exposure (less than 30 minutes in the case of some studies) in a single laboratory training session (Dulany et al., 1984; Leow & Hama, 2013). Following this line of argument, Serafini (2013) has suggested an alternative interpretation of the data that source attributions provide—specifically, that the data from this measure might be best interpreted in terms of degrees of explicitness, rather than necessarily reflecting implicit or explicit knowledge as binary alternatives. In other words, when using *guess*, *intuition*, *memory*, or *rule* as the possible attributions, *guess* and *intuition* would reflect no or a relatively low level of awareness, whereas *memory* or *rule* might reflect a higher level of awareness. This interpretation is supported by research in cognitive psychology that has examined the time course of the development of implicit and explicit knowledge (Ziori & Dienes, 2012), as well as by several studies in SLA that have utilised concurrent measures of awareness, such as think-aloud protocols (Leow, 1997; Rosa & Leow, 2004). These studies have reported that awareness is not a dichotomous construct, as the results of retrospective verbal reports tend to suggest (see Brooks & Kempe, 2013 for an exception), but instead operates on a continuum from no awareness to awareness.

3.5.4. Direct Versus Indirect Tests

Following initial proposals by Reingold and Merikle (1988), the use of direct and indirect tests to distinguish between implicit and explicit knowledge has received support from several researchers within cognitive psychology and SLA (see Rebuschat, 2013 for an overview). This support within SLA is perhaps unsurprising, given the fact that task bias, or that some tasks promote the use of implicit knowledge and other tasks promote the use of explicit knowledge, has long been a point of discussion (Andringa & Rebuschat, 2015; Bowles, 2011; Doughty, 2003; Han & R. Ellis, 1998; Norris & Ortega, 2000). Broadly defined, direct tests ask participants to make direct use of their knowledge, and are argued to be reflective of explicit knowledge. By contrast, indirect tests, which are held to be more accurate measures of implicit knowledge, gauge knowledge indirectly, such as through behavioural data; for example, differences in reaction times when participants are confronted with regular versus irregular stimuli. An example of a direct test is a traditional grammaticality judgement test; an example of an indirect test is the self-paced reading test (both of which are discussed in Section 2.6). Within SLA, the best-known examples of direct and indirect tests stem from the recent psychometric studies of R. Ellis and colleagues (2004, 2005; Ellis et al., 2009; Godfroid et al., 2015; see also Isemonger, 2007; Zhang, 2015). The initial studies in this series (Ellis, 2005) concluded that untimed GJTs and metalinguistic knowledge tests are better measures of explicit knowledge, whereas timed GJTs, elicited imitation tasks, and oral narration tasks draw primarily upon implicit knowledge. Other SLA researchers (Leung & Williams, 2011, 2012, 2014; Paciorek & Williams, 2015a, 2015b) have drawn upon the SRT methodology and have utilised reaction time measurements as indirect measures of implicit knowledge.

Although direct and indirect tests have been argued to be more sensitive than retrospective verbal reports in disentangling implicit and explicit knowledge (Reingold & Merikle, 1990; Shanks & St. John, 1994), these tests are not without their limitations. Firstly, as noted in Section 2.4.6, it is generally well accepted in the literature that learners draw upon both implicit and explicit knowledge to complete any given task (DeKeyser, 2009; N. Ellis, 2005), and that the characteristics of the task itself influence the degree to which learners draw more heavily upon either implicit or explicit knowledge (Bialystok, 1979, 1982, 1986). In other words, certain tasks can encourage the use of either implicit or explicit knowledge in L2 performance, but it would be erroneous to conclude that a particular task is reflective of purely implicit knowledge or purely explicit knowledge. To illustrate, untimed GJTs are argued to be reflective of explicit knowledge, and timed GJTs are argued to be reflective of implicit knowledge (R. Ellis et al., 2009). This distinction is predicated on arguments that implicit knowledge is more available for rapid, automatic processing, whereas explicit knowledge is available for slower, more controlled processing (see Section 2.3.3). Thus, it is believed that time pressure encourages the use of implicit knowledge, because participants do not have time to draw upon their explicit knowledge to complete the task. There are, however, several potential problems with this interpretation. Firstly, with regard to the untimed GJT, which is argued to be more reflective of explicit knowledge, there is no guarantee that participants are not drawing upon their implicit knowledge to complete the task. In other words, this result from an untimed GJT could be contaminated by implicit knowledge (Reingold & Merikle, 1988; see also Rebuschat, 2013). In addition, the results of a timed GJT, which is argued to reflect implicit knowledge, could be contaminated by automatised explicit knowledge (DeKeyser, 2009).

Finally, from a practical viewpoint, the amount of time provided for participants is problematic. This issue arises from the challenge of finding the "sweet spot" that allows participants to process the sentence semantically while limiting the time available for them to draw upon their explicit knowledge (R. Ellis et al., 2009). If the time limit for items on the GJT is arbitrarily assigned[9], then there is no guarantee that the imposed time limit actually influences the manner in which the participants process the stimuli. Some studies, however, have gone to considerable lengths to establish the validity and reliability of the time limits used in their instruments. In Loewen's (2009) study, the test items were first trialled on native speakers of English who were encouraged to answer the questions as quickly as possible. The median response time was calculated for each item; then, due to the fact that L2 learners would be taking the test, an extra 20% was added to each item. This resulted in varying time limits for different items, which ranged from 1.8 seconds to 6.24 seconds (Loewen, 2009). There are a number of issues here, which are not limited to the use of L1 norms in SLA research. Firstly, although these reaction times were later validated psychometrically, it is not clear whether the results here can be generalised beyond the participants (both L1 and L2) and the test items included in this particular sample. The implications of this are that SLA researchers should not

[9] A limitation of research that has utilised timed GJTs (Han, 2000) is that these studies often provide no information about how they established the cut-off times as part of their instruments.

expect that the same time limits would necessarily be valid for a different population of learners, or even for the same population of learners with different test items.[10]

3.5.5. Triangulating Measures of Awareness

The measurement of awareness has long been a contentious issue in the debate regarding implicit and explicit processes in SLA. This debate is partly due to the slippery nature of implicit and explicit knowledge, and partly to issues related to the instruments researchers have used to distinguish between implicit and explicit knowledge (Andringa & Rebuschat, 2015; DeKeyser, 2003). One issue related to instrumentation is that, when relying on a single method for collecting data, it is often difficult to confirm that the observed patterns in the elicited data are not simply an artefact of the method itself (Chaudron, 2003; Purpura, Brown, & Schoonen, 2015). As such, to arrive at more valid and reliable findings, researchers investigating implicit and explicit processes in SLA have increasingly begun to triangulate data by employing multiple measures of awareness. This progression can be readily seen in studies that utilised a single concurrent (Leow, 2000) or retrospective (Williams, 2005) measure of awareness, to studies that have incorporated two measures, such as subjective measures in addition to retrospective verbal reports (Hamrick & Rebuschat, 2013; Rebuschat & Williams, 2012), indirect measures in addition to retrospective verbal reports (Leung & Williams, 2011, 2012, 2014), and concurrent measures, such as think-aloud protocols, in addition to retrospective verbal reports

[10] In addition, it might be more valid to use an untimed GJT, but to ask participants to make their decisions as quickly as possible. Studies such as those included in this thesis (see also Rebuschat & Williams, 2012) have reported that participants responded within 2000 ms, which would fall at the low end of the range utilised by R. Ellis et al. (2009). Answers with shorter/longer reaction times could be compared using subjective measures of awareness. Such an approach would obviate the need for a lengthy validation process in setting time limits for individual items on a GJT.

(Hama & Leow, 2010). More recently, researchers have taken this a step further. For example, Rebuschat et al. (2015) triangulated data from think-aloud protocols, retrospective verbal reports, and subjective measures of awareness. The results of this study confirmed previous findings of implicit knowledge (Rebuschat & Williams, 2012), provided further evidence of the validity and applicability of subjective measures of awareness to SLA research and, most interestingly, found evidence of reactivity; in other words, that the measures of awareness may impact on the measurement of learning. Another recent study by Godfroid et al. (2015) used eye-tracking in conjunction with direct and indirect tests. The results of this study were argued to confirm that these tasks measured different types of linguistic knowledge[11] (R. Ellis, 2005; R. Ellis et al., 2009; Zhang, 2015).

3.6. Summary

Rebuschat and Williams (2012) and the other studies listed above (such as those by R. Ellis et al., 2009; Godfroid et al., 2015; Hamrick, 2013; Rebuschat et al., 2013; Rebuschat et al., 2015; Serafini, 2013; Williams, 2005 and Zhang, 2015) are indicative of the renewed attention in SLA towards the incidental learning of L2 grammar, the type of knowledge that is developed over the course of incidental exposure, and methodological issues in the measurement of this knowledge. As can be gathered from the discussion, the measurement of implicit and explicit knowledge has proved both elusive and controversial, and more research into the effectiveness of measures of awareness is clearly warranted. One limitation of the extant research is that the vast majority of these studies have addressed the incidental learning of L2

[11] Please refer to Section 3.5.4 for a critique of this claim.

word order (Rebuschat & Williams, 2012; Williams & Kuribara, 2008) or noun-determiner systems (Hama & Leow 2010; Leung & Williams, 2011, 2012; Rebuschat et al., 2013; Williams, 2005). To date, very little research has investigated the learning of L2 inflectional morphology following incidental exposure with the aim of examining whether the resulting knowledge was implicit or explicit in nature (for exceptions, see Grey et al., 2014 and Marsden et al., 2013). The next chapter provides a brief review of the research that has examined the acquisition of L2 morphology, which is the focus of the present study.

CHAPTER 4

PROCESSING AND ACQUISITION OF L2 MORPHOLOGY

Chapter 4 of this thesis provides the background to the incidental learning of L2 inflectional morphology, and the role of awareness in this process. As issues related to processing and acquisition are intertwined, this chapter begins with an overview of two influential models of how learners process input in their L2: VanPatten's model of input processing (VanPatten, 2002a), and the shallow structure hypothesis (Clahsen & Felser, 2006). Following this, the chapter shifts focus and reviews the relevant SLA literature regarding the acquisition of L2 case markings. This chapter provides a detailed analysis of three studies that have investigated the role of awareness in the learning of L2 case markers: Brooks and Kempe (2013), Grey et al., (2014), and an early study by Robinson (2002, 2005). Finally, the chapter concludes with the research questions of this thesis.

4.1. Factors Influencing the Processing and Acquisition of L2 Morphology

There has been long-standing interest in the acquisition of L2 morphology within the field of SLA. The sustained importance and relevance of the issue is well reflected in the fact that two journals, *Language Learning* and *Language Teaching Research*, have devoted special issues to research on L2 morphology in recent years. This continued interest can partly be attributed to the large number of published studies that have demonstrated that L2 learners process inflectional morphology differently from native speakers (Clahsen, Felser, Neubauer, Sato & Silva, 2010; Jiang, 2004, 2007), and face considerable difficulty in acquiring both receptive and productive knowledge of L2 morphology (DeKeyser, 2005; Hopp, 2013; Larsen-Freeman, 2010; White, 2003, 2009). Although this body of research has identified a wide variety of factors that impact the degree to which learners are successfully able to acquire morphological forms (see Section 2.9.3.3), issues related to attention and awareness have been strongly implicated in the success of learners in this endeavour (Brooks & Kempe, 2013; N. Ellis, 2006a, 2006b). Of particular importance to the debate regarding the role of attention and awareness in the acquisition of L2 morphology is the synergistic relationship between L2 processing and L2 acquisition (Andringa & Curcic, 2015; Clahsen & Felser, 2006; VanPatten, 2002a). Of particular concern is the fact that learners exhibit what has been referred to as a "default processing bias" (Park & Han, 2008, p. 109), meaning that they tend to divert more attentional resources to particular elements of a sentence, such as content words, than they do others, such as morphological forms (see Clahsen & Felser, 2006; Sagarra, 2008 and VanPatten, 2014 for examples). This processing bias is theorised to impact on the degree to which learners are able to acquire these forms successfully. In the next sections, I will briefly define additional terms related to the processing and acquisition of L2 morphology before outlining relevant models of input processing from SLA and psycholinguistics.

4.2. Input Processing: Key Terms and Definitions

It is important to define further terms related to the stages of the acquisition process. Firstly, it is well established that learners do not incorporate all aspects of what they read or hear into their interlanguage system. In this regard, we can distinguish between the concepts of *input* and *intake* (Corder, 1967). Input in the second language acquisition literature refers to the linguistic information that the learner is exposed to in either oral or written form. Intake, however, represents the subset of input utilised in some manner by the learner prior to any learning taking place (Gass, 1997; Leow, 2012, 2015b; Slobin, 1985; VanPatten, 2004b, 2007). As can be seen, intake then represents an intermediary step between the raw linguistic data that are available to the learner, or input, and the incorporation of a subset of these data into the learner's interlanguage system.

When we consider that intake represents only a subset of the data available in the input, a question arises about the process and/or mechanism that mediates the aspects of the input that become intake and the aspects that do not. Input processing in the SLA literature is theorised to take place between the input and intake stages outlined above (Gass, 1997; Leow, 2012, 2015b; Robinson, 2003; Schmidt, 2001; VanPatten, 2004b, 2007). Processing refers to mental processes through which aspects of the input, whether meaning or form, are selectively attended to and made available for further processing (VanPatten, 2004b).

4.3. Models of Input Processing

4.3.1. VanPatten's Input Processing Model

Perhaps the most widely discussed model of input processing in SLA is VanPatten's model of input processing (1996, 2002a, 2002b; 2004b; 2005; 2007, 2015b). As noted in the introduction to Section 4.1, learners display biases towards the aspects of the input to which they naturally attend. VanPatten's input processing model attempts to account for this "processing bias" by providing a theoretical framework for the default manner in which L2 learners make their initial form-meaning connections as they select, attend to[12], and process input in an L2—in other words, the manner in which certain aspects of the input become intake (VanPatten, 2002a; Sanz & VanPatten, 1998). The central claims of VanPatten's input processing model are that

a) learners are driven towards meaning when attempting to comprehend input,

b) the comprehension of input is a cognitively demanding process, particularly in the early stages of language acquisition (VanPatten, 2007, 2015a, 2015b), and

c) the degree to which learners attend to grammatical form is influenced by a number of guiding principles (Park & Han, 2008; Sagarra, 2008; VanPatten, 2004b).

Although a full discussion of VanPatten's model is beyond the scope and focus of this thesis, several of these principles appear to be both useful and relevant for understanding the development of knowledge of L2 case marking. These principles include the primacy of meaning principle and the lexical preference principle (for discussions of VanPatten's model, see Benati, 2013; Lee, 2015; Lee & Benati, 2013; Sharwood Smith, 2015; VanPatten, 1996, 2004b, 2007, 2015b; Wong, 2004).

As noted, VanPatten's input processing model has stipulated a number of principles that guide how learners initially process input in the L2. What may be seen as the central principle in this account is referred to as the primacy of meaning principle, which holds that learners will first and foremost devote attentional resources to comprehending the meaning of an utterance (VanPatten, 2004b). This

[12]The constructs of processing and noticing are not isomorphic in this framework (see VanPatten, 2007, 2015b). The processing of information goes beyond merely noticing features of the input (Leow, 2015a) because it involves the learner making the connection between form and meaning. Although noticing can be a component of processing (Han & Sun, 2014; VanPatten, 2014), input processing can proceed in the absence of awareness of what is being processed (N. Ellis, 2002; see VanPatten, 2015b for a discussion).

principle is predicated on research into working memory capacity (Baddeley, 1986, 2000, 2007, 2012; Baddeley & Hitch, 1976), particularly with regard to the concept of working memory as a limited capacity system that influences lexical and syntactic processing in an L2 (Just & Carpenter, 1992; Miyake & Friedman, 1998; see Juffs, 2004, 2005; Juffs & Harrington, 2011; Wen, 2012; Williams, 2012 for overviews). In this view, a finite amount of cognitive resources is available for any given task. For learners to devote attentional resources to the processing of some grammatical form, such as redundant inflectional markers, the processing of the meaning of a sentence must not drain the available cognitive resources.

Another of VanPatten's (2004b) input processing principles that appears to be particularly relevant to the difficulties learners face when processing L2 morphology is the lexical preference principle, which holds that learners tend to derive the meaning of a sentence from lexical items, particularly from content words, as opposed to from grammatical forms. Furthermore, this principle is important with regard to the processing of morphological forms, in that "if grammatical forms express a meaning that can also be encoded lexically (i.e., that a grammatical marker is redundant), then learners will not initially process those grammatical forms" (VanPatten, 2007, p. 118). An example of this principle might be how a learner processes a sentence such as "Last night I walked on the beach". In this case, the past marker –*ed* is redundant because past meaning is also encoded lexically through the phrase "last night". In this case, input processing holds that learners would process the redundant marker –*ed* at a later stage than at which they would process "last night", provided that sufficient cognitive resources are available for this task.

It is important to reiterate that input processing, as set out by VanPatten (2004b, 2007, 2015a), does not represent an overarching theory of second language acquisition, but is instead a model of how learners make the initial connections between meaning and form as they attempt to make sense of L2 input (VanPatten, 2007). Although many have expressed concerns regarding the soundness of VanPatten's model on a theoretical level (Carroll, 2004; DeKeyser, Salaberry, Robinson, & Harrington, 2002; Harrington, 2004), several studies have provided both direct and indirect support for many of VanPatten's claims (Cox & Sanz, 2015; Han & Sun, 2014; Marsden, 2006; Marsden & Chen, 2011[13]; Park, 2014; Sagarra, 2008; Sagarra & N. Ellis, 2013; Sanz & Morgan-Short, 2005; Sanz & VanPatten, 1998; VanPatten & Cadierno, 1993). For example, in relation to principles connected with redundancy, research investigating the online processing of verbal morphology has provided evidence that learners might not process redundant forms at all (Sagarra, 2008; Sagarra & N. Ellis, 2013). Further indirect evidence can be found in research investigating the acquisition of L2 grammar that has identified redundancy as a mitigating factor in the acquisition of these forms (Bardovi-Harlig, 2000, 2007; N. Ellis, 2002, 2003, 2007a; Gass, 1997; Robinson, 2002, 2005, 2010).

Despite finding some support in the SLA literature, VanPatten's model of input processing is not without its detractors, particularly regarding the degree to which VanPatten's model aligns with the greater body of research on input processing

[13] It is important to note that the results of these studies do not provide unwavering support for many of VanPatten’s claims. See, for instance, the discussion related to the development of automatized explicit knowledge in Marsden and Chen (2011).

within psycholinguistics (Carroll, 2004; DeKeyser et al., 2002; Harrington, 2004). Criticisms of VanPatten's model have included the fact that

a) this model is predicated on antiquated models of attention (DeKeyser et al., 2002),

b) some of the guiding principles of VanPatten's model, such as the sentence location principle, contradict much of what is known about how learners parse syntactic structure (Carroll, 2004),

c) VanPatten's model makes no acknowledge of the role of the L1 in L2 processing (Carroll, 2004),

d) this model makes no provisions for the role of explicit knowledge in L2 processing (Marsden, 2006), and

e) VanPatten's model does not actually address how input is processed, but rather how input is comprehended (Harrington, 2004).

4.3.2. Psycholinguistic Models of Input Processing

Among psycholinguists, the term *input processing* has a comparatively broad definition in that it can refer to:

> ...any dynamic operation in real time that converts a stimulus into a message or a message into a motor-articulatory plan. Among psycholinguists investigating speech perception and sentence comprehension, the term processing can refer to any dynamic operation in real time that converts a stimulus into a message or a message into a motor-articulatory plan. Processing from this perspective is stage-like and includes everything from the subconscious detection of phonetic distinctions relevant to word recognition, through morphosyntactic parsing, and message integration (Carroll, 2004, p. 294).

Research in this area has utilised a wide range of experimental methods, such as event-related potentials, self-paced reading, and semantic priming as part of lexical decision tasks (for overviews, see Kaiser, 2013; Keating & Jegerski, 2015; Morgan-

Short & Tanner, 2014; Roberts, 2012a) to examine the mechanisms and limitations that native speakers and language learners employ as they process word-level and sentence-level information in real time. This research is concerned with elucidating how learners construct linguistic representations during language comprehension and production (Clahsen et al., 2010).

Although processing research is a broad area of enquiry, what is most relevant to this thesis are subdomains of processing research that have focused on sentence, or syntactic processing, and studies investigating morphological decomposition. Sentence processing, or syntactic processing, is concerned with how learners parse syntactic relationships within a sentence in real time. Here, the term *parsing* refers to how learners compute syntactic information incrementally in real time. VanPatten and Benati (2010) provide an example:

> ...the moment a listener hears "the man...," that listener immediately projects a determiner phrase (DP) while simultaneously tagging it as "subject of sentence". If the listener hears "reduced" next, then that listener most likely tags the word as "verb," "past tense," thus projecting a verb phrase and confirms that "the man" is the subject of the verb. However, if the listener next encounters "to tears," then the listener's parsing mechanism stops and reanalyses "the man reduced to tears..." as a reduced relative clause that is the subject of a sentence....And so the analysis progresses as each word is encountered (p. 123)

Hence, parsing involves a real-time process of assigning syntactic roles to various sentence components and making projections and predictions about syntactic relationships to come in an attempt to make sense of the utterance. In contrast to parsing, which investigates the manner and extent to which learners process syntactic relationships within a sentence, other research (Clahsen et al., 2010) has focused on morphological decomposition, or the manner in which the learner processes individual words. The chief issue in this line of study is whether individual words are processed and stored as whole-word representations, or whether individual words are "decomposed" - broken down, processed and stored - according to their constituent parts.

Recent research investigating syntactic processing (Bialystok & Craik, 2010; Dussias, 2003, 2010; Dussias & Sagarra, 2007; Felser, 2012; Felser & Clahsen, 2009; Felser, Roberts, Marinis & Gross, 2003; Fernández, 2003; Frenck-Mestre, 2002; Frenck-Mestre & Pynte, 1997, 2000; Jacob & Felser, in press; Kielar, Meltzer, Moreno, Bialystok, & Alain, 2014; Moreno, Bialystok, Wodniecka, & Alain, 2010; Osterhout, McLaughlin, Pitkanen, Frenck-Mestre, & Molarino, 2006; Roberts, 2012b; Roberts & Felser, 2011; Roberts & Liszka, 2013) and morphological decomposition (Clahsen et al., 2010; Clahsen & Neubauer, 2010; Gor & Jackson, 2013; Jiang, 2004, 2007; Kirkici & Clahsen, 2013; Neubauer & Clahsen, 2009; Silva & Clahsen, 2008) has pointed to marked differences in the manner in which native speakers and adult second language learners process incoming linguistic information. In general, this line of research has indicated that L2 learners are less sensitive to both syntax (Roberts & Felser, 2011) and morphology (Jiang, 2004), and rely more heavily on semantic and lexical clues than on the grammatical information encoded in these forms.

Several models have been proposed to document the differences in processing behaviour between L1 speakers and adult L2 learners[14]. Central to the theoretical discussion of morphological processing is whether learners employ a single- or dual-

[14] Since the focus of this thesis is the acquisition of L2 inflectional morphology, the discussion will be limited to how these theoretical models address morphological processing. For discussions of how these and other models account for syntactic processing, see Carroll (2001, 2007), Clahsen & Felser (2006), and Harrington (2010).

mechanism system to process regular and irregular inflectional forms. To put it simply, the single-mechanism account maintains that all inflected words, both regular and irregular forms, are processed in the same manner of associative patterning (Bybee, 1995; Seidenberg & Elman, 1999). By contrast, the dual-mechanism view (Pinker, 1999; Pinker & Ullman, 2002) maintains that regular verbs are decomposed and processed by symbolic rule computation, whereas irregular verbs are retrieved whole via associative memory.

Which model would better fit adult SLA? It has been argued that it is "incontestable" that the beginning stages of L2 learning involve the rote memorisation of unanalysed chunks of language (N. Ellis, 2001; Gor & Jackson, 2013), and that in order to pass beyond the beginning stages of language acquisition, learners need to begin to process and acquire a wide number of inflected words "for which they do not have stored whole-word representations" (Gor & Jackson, 2013, p. 1066). As such, some form of a dual-system account, in which learners rely on both rote memorisation and increased decomposition at higher levels of proficiency, appears to be a better fit with regard to L2 learning.

Two models of morphological processing, the declarative/procedural model (Ullman, 2001, 2004, 2013) and the shallow structure hypothesis (Clahsen & Felser, 2006), maintain a dual-system view of morphological processing in that they hold that native speakers decompose regularly inflected words, whereas irregular words are stored as a whole in associative memory. The declarative/procedural model posits that L2 learners rely more heavily on declarative memory and store regular words whole in the beginning stages of language acquisition. It is only at higher stages of proficiency that learners begin to decompose words and start to utilise procedural memory as they process morphological forms. By contrast, the shallow structure hypothesis holds that adult L2 learners process input in a shallower manner than do adult native speakers (see VanPatten, 2014, for a concise overview). The basic claim of this model is that input processing can take two different forms. The first is structural processing, in which syntactic structure is fully processed during comprehension (as per the example of parsing above). The second is shallow processing, in which syntactic structure is only partly processed during comprehension, and entails greater reliance on lexical information. While native speakers utilise both structural and shallow processing, Clahsen and Felser argued that second language learners only engage in shallow processing (Clahsen & Felser, 2006). In its initial form, Clahsen and Felser (2006) proposed that marked differences between how adult native speakers and adult L2 learners process syntactic relationships can clearly be seen; specifically, that L2 learners rely on lexical and other non-structural clues in their processing of an utterance. In contrast to the processing of syntactic structure, Clahsen and Felser (2006) argued that adult L2 learners appear to process simple morphological rules in a native-like fashion. However, in a follow-up study, Clahsen et al. (2010) found that adult L2 learners were less sensitive to morphological inflections than were native speakers, and tended to rely on lexical-semantic relationships, rather than on morphological decomposition in their processing of L2 input. This result has been corroborated by a more recent series of experiments (Krause, Bosch, & Clahsen, in press) that has indicated that both lexical representation and processing by non-native speakers relies less on morphosyntactic information than is the case for native speakers. In summary, these more recent follow-up studies support the argument that non-native speakers' processing of both syntax and morphology is "shallow" in the sense that they rely less on morphosyntactic clues than do native speakers.

4.3.3. Processing of L2 Morphology

Despite the evidence that L2 learners at all proficiency levels appear to process L2 input differently than do native speakers, research to date and the theoretical models described above point towards a symbiotic relationship between grammatical competence and the degree to which these grammatical forms are processed in the input (Andringa & Curcic, 2015; VanPatten, 2014). In other words, the degree to which L2 learners process relevant grammatical forms appears to influence the degree to which these forms are acquired, although it is important to note that the exact nature of this relationship remains unclear (Mitchell et al., 2013). Conversely, overall proficiency level and grammatical knowledge also directly impact on L2 processing. It should, however, be stressed that the extent to which L2 learners process morphological markers in an L2 cannot simply be accounted for by their overall L2 proficiency. Research investigating the acquisition of L2 morphology has pointed to several additional factors that influence the processing and acquisition of grammatical morphemes. These factors include the type and token frequency of the target inflection in the input (N. Ellis, 2002, 2003; N. Ellis & Schmidt, 1997, 1998), L1 background (N. Ellis & Sagarra, 2010a, 2010b, 2011; Rast, 2010; Vainio, Pajunen, & Hyönä, 2014), individual difference variables (Brooks et al., 2006; Kempe & Brooks, 2008; Kempe et al., 2010; Sagarra, 2008), metalinguistic awareness of target morphological constructions (Brooks & Kempe, 2013; Grey et al., 2014; Marsden et al., 2013; Robinson, 2005), and type of instructional condition (Alanen, 1995; DeKeyser, 1995; Marsden et al., 2013; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010).

Given the wide range of factors identified, several researchers have attempted to synthesise this body of research and to narrow down the variables that influence the processing and acquisition of L2 morphological forms. For example, drawing on research investigating the order of acquisition in English SLA, Goldschneider and DeKeyser (2001) identified five determinants that account for the variance in the acquisition of grammatical morphemes: perceptual salience (how easy it is to perceive a given structure), semantic complexity (how many meanings are expressed by a particular form), morphophonological regularity (the regularity with which a morpheme is expressed in different environments; for example, allomorphy or a particular morpheme), syntactic category (lexical versus syntactic morphemes, free versus bound), and frequency in the input. Further linguistic and extralinguistic factors with regard to the processing of L2 morphology were identified by Gor (2010), including the richness, complexity and predictability of inflectional paradigms in the L2 (decomposability), properties of the L1 (whether inflections exist in the learner's L1; L1 transfer)[15], L2 proficiency, properties of the input, amount of exposure, and individual differences (such as verbal working memory capacity).

[15] See also Jarvis (2000) and Jarvis and Odlin (2000) for a discussion on the effects of the L1 on the acquisition of additional languages.

4.3.4. Input processing: Summary

In summary, the processing and learning of L2 morphology is an area that has attracted considerable attention within SLA. Research in this area has testified that the learning of this area of grammar is not a straightforward task, but is influenced by a wide range of interrelated factors connected with attention and awareness, including the L1 and L2 of the learner, and individual difference variables, such as verbal working memory ability, which is unique to the learner in question. This research has included studies looking directly at L2 processing (Sagarra, 2008) and/or L2 acquisition (Marsden et al., 2013), as well as various incarnations of morphology, such as verbal versus nominal affixation, and free versus bound morphological forms. Because this present thesis is directly concerned with the acquisition of L2 case-marking systems, the next section will narrow its focus accordingly and will examine studies that have investigated the acquisition of L2 case markers directly, with particular attention being paid to empirical research that has set out with the aim of accounting for the role of awareness in this process.

4.4. Learning L2 Case-marking Systems

The majority of research that has investigated the learning of L2 case marking has utilised instructed or intentional learning conditions as part of its research design (Brooks, Kempe, & Donachie, 2011; Hinz, Krause, Rast, Shoemaker, & Watorek, 2013; Kempe & Brooks, 2008; Kempe et al., 2010; Brooks et al., 2006; Kempe & MacWhinney, 1998; Rast, 2008, 2010; Rast, Watorek, Hilton & Shoemaker, 2014). Thus far, very little research has examined the degree to which L2 case marking can be acquired under incidental conditions, particularly after a minimal amount of exposure, and even fewer studies have set out to examine the role of awareness in this process. To my knowledge, this gap in the literature, specifically concerning the role of awareness in learning L2 case marking, has been addressed by only three studies to date, two of which investigated this issue under incidental learning conditions (Grey, Williams, & Rebuschat, 2014, 2015; Robinson, 2002, 2005), and a single study reported in two separate articles under intentional learning conditions (Brooks & Kempe, 2013; Brooks et al., 2011).

4.4.1. Grey, Williams, and Rebuschat (2014, 2015)

Grey et al. (2014) adapted the artificial language developed by Williams and Kuribara (2008, see Section 2.7.1.2) to examine the acquisition of both L2 word order and case marking. As did Williams and Kuribara (2008), Grey et al. used “Japlish,” a semi-artificial language consisting of English words and Japanese syntax, in which English words were inflected according to their function in the sentence (for example, *Stacey-ga that picture-o painted*). In the training phase, participants in Grey et al.'s study were exposed to 128 Japlish sentences auditorily, and they were asked to judge the semantic plausibility of each sentence. In Williams and Kuribara’s original (2008) study, the testing phase consisted solely of a GJT that assessed the learning of L2 word order. Grey et al. modified the testing phase of the study to include both a GJT, designed to test the learning of L2 word order, and a picture-matching task that assessed the learning of the morphological case markers. In the picture-matching task, participants had to indicate if a sentence such as *Man-o woman-ga necklace-ni gave* matched the picture displayed on a monitor; in this case, a man giving a woman a necklace. The testing phase of this study was further modified to include both immediate and delayed post-tests. In addition, confidence ratings and source attributions were taken alongside the grammaticality judgement test, but not the picture-matching task. This was followed by retrospective reports at the end of the testing phase.

The results of the GJT showed a significant learning effect of word order on both immediate and delayed post-tests. An analysis of confidence ratings and source attributions from the GJT indicated that participants were most accurate when basing their decisions on intuition, rather than on explicit knowledge of the L2 syntax. In other words, the participants developed some implicit knowledge of the L2 word order as a result of incidental exposure. With regard to the acquisition of case marking, the results from the picture-matching task showed a significant learning effect only for case-marking on the delayed post-test. In addition, an analysis of the retrospective verbal reports indicated that the overall learning effect of L2 case markings was driven by participants who expressed awareness of the morphological system. This suggests that the knowledge of L2 case markings acquired in this experiment was largely explicit in nature. However, since awareness was only measured retrospectively, it is impossible to pinpoint when the participants became aware of this regularity. It is possible that the participants might have become aware of the morphological system during the testing phase of the experiment. In addition, because subjective measures of awareness were only obtained for the grammaticality judgement test, it is not possible to say whether the participants were partly basing their decisions on intuition, rather than solely on explicit knowledge of the underlying morphological rule system.

In summary, the results of Grey et al. (2014) indicate that it is possible to acquire some knowledge of a novel morphological system following incidental exposure. There were, however, a few limitations to this study in terms of the measurement of awareness that were not addressed by the authors. Firstly, as mentioned above, although subjective measures of awareness were taken alongside the GJT, the picture-matching task was not accompanied by such measures. As a result, any conclusions regarding the nature of learning with regard to case markers can only be inferred based on the results of the offline verbal reports, thus jeopardising the construct validity of the findings. Secondly, the design of the picture-matching task did not allow for analysis of performance on individual morphological markers. For example, in the testing phase, participants were asked to indicate if the sentence presented matched the picture displayed on the screen. Because the stimulus sentences for this test, such as *Man-o woman-ga necklace-ni gave,* contained both nominative and objective case markers, it was not possible to distinguish the quality of the participants' performance on the two morphological markers. The learning effect on one of the case endings might have been masked by poor performance on the other case ending.

4.4.2. Robinson (2002, 2005, 2010).

In contrast to Grey et al. (2014), an earlier study by Robinson (2002, 2005, 2010) did allow for comparisons across isolated morphological case markers. In Robinson (2005), a miniature Samoan system was used to examine the acquisition of L2 morphology, specifically locative, ergative and noun incorporation case markers. First, the participants, L1 speakers of Japanese, were trained to criterion on translation equivalents of the Samoan lexis that would be used in the next phase of the study. Next, a Samoan sentence was presented to the participants on a monitor for 10 seconds. Based on the content of this sentence, participants had to answer a yes/no comprehension question and were given feedback after each answer. This training included ten trials consisting of 45 sentences each. Thus, participants were exposed to a total of 450 target structures. Finally, the participants completed immediate, one-week and six-month delayed written and auditory grammaticality judgement tests, guided sentence production tests, and a retrospective written awareness questionnaire. The results of this study indicated a significant learning effect only for the locative markers on the immediate and delayed post-tests. As in Grey et al. (2014), however, the results of the retrospective questionnaire indicated that participants were largely able to verbalise the rule governing the use of the locative marker, suggesting that the overall learning effect was driven by explicit knowledge. Robinson (2005) noted that the results here suggested that incidental learning might not be truly "incidental." In other words, incidental learning is not wholly an unconscious and implicit process, as assumed by previous AGL research (Reber, 1967), but that incidental learning encompasses both conscious and unconscious cognitive processes (N. Ellis, 2001, 2005).

4.4.3. Brooks, Kempe, and Donachie (2011).

Finally, a study by Brooks et al. (2011) investigated the learning of gender categories, case markings, and lexis of a miniature Russian system. This study followed what is known as a micro-genetic design (Sielger & Crowley, 1991). In contrast to the AGL paradigm, which typically consists of a single, short training session, micro-genetic studies typically employ multiple training sessions or trials over an extended time frame. Building on their previous research, which examined the role of diminutives in the acquisition of Russian gender marking (Kempe & Brooks, 2001; see Section 2.7.1), this study examined the role of diminutives in the acquisition of gender categories (noun-adjective agreement), case markings (genitive and dative cases), and Russian lexical items. Unlike their previous experiment (Kempe & Brooks, 2001), all participants in this study were exposed to both simplex and diminutive nouns of a miniature Russian system. The training and testing sets consisted of a total of 24 nouns (12 masculine, 12 feminine). Each training noun was presented in two of three contexts (dative, genitive, noun-adjective agreement) in the training phase. The remaining context was used in the testing set to assess the extent to which participants could generalise to trained nouns. The eight nouns not included in the training set were presented in all three contexts in the testing set. In addition, all nouns in this study were transparent in terms of their respective genders—all feminine nouns ended in *-a*, and all masculine nouns ended in a consonant. In the diminutive forms, feminine nouns ended in *-ka*, and masculine nouns ended in *-ik*.

The training phase of this study consisted of six one-hour training sessions carried out over a period of 14 days. All the training sessions consisted of four blocks, each of which required the participants to perform different tasks. In Block 1 (listen and repeat), participants were presented with a line drawing while listening to a dialogue in Russian between a man and a woman. For example, in the dialogue, the participants would hear a male voice say "*Otkuda ukhodit slon*?" (From where is the elephant coming?). A female voice would answer, "*ot chajnika*" (From the kettle).

While listening to this dialogue, a picture of an elephant walking away from a kettle would appear on the computer screen. The participants' task was to repeat the phrase spoken by the female voice. The recording was then played again, and the participants were asked to repeat the phrase a second time. This process was repeated for a total of 48 items—16 items corresponding to the genitive case, 16 to the dative case, and 16 connected with adjective-noun gender agreement.

In Block 2, participants were presented with two pictures. They then listened to a recording of a noun, such as "*chajnik*" [kettle], and indicated the picture that matched the recording. After the participants had made their choices, the noun was repeated and the participants were provided with feedback by being shown the correct picture on the monitor. Block 3 consisted of a case-comprehension task in which participants listened to the same dialogues that were played in Block 1. While the dialogues played, two pictures were presented on the screen, both showing the same object in different contexts for case marking (for example, an elephant walking towards or away from a fence) or the same object in different colours for noun-adjective agreement. Participants listened to the dialogue, then chose the correct picture. After the participants made a choice, the dialogue was repeated, and they were provided with feedback by being shown the correct picture on the monitor. Finally, Block 4 was the production block. Participants were shown a picture of, for example, an elephant walking away from a kettle. They were provided with the name of the kettle in the nominative form. They then listened to the first half of the dialogues from Blocks 1 and 3. In this case, they would hear the male voice ask

"*Otkuda ukhodit slon*?" (From where is the elephant coming?). They would then need to answer this question in Russian.

The testing phase of this study consisted of three parts. Part 1 of the testing phase assessed the degree to which the participants had acquired the case marking and gender agreement of the miniature system. This part was identical to Block 4 of the training phase and consisted of 96 trials, 48 trained items and 48 transfer items. Unlike in the training phase, no feedback was provided. In Part 2, participants were tested on the degree to which they had learned the Russian lexis used in the experiment. Participants were presented with line drawings one at a time and were asked to name each one. The final part of the testing phase consisted of a battery of tests assessing individual difference variables, and a debriefing questionnaire. The results of Part 3 are discussed in Brooks and Kempe (2013) below.

The results of this study indicated a high degree of variance in the performance of the participants. Overall, participants were more accurate on trained items (77% accuracy) than they were on transfer items (42% accuracy) for case-marking items. In addition, participants were more accurate with nouns presented in the diminutive form (65% accuracy) versus the simplex form (55% accuracy), and for masculine nouns (66% accuracy) compared with feminine nouns (53% accuracy). In terms of the vocabulary test, the participants were more accurate in recalling diminutive nouns versus simplex nouns, and masculine nouns over feminine nouns. Overall, the participants were more accurate when generalising case marking and gender with diminutive forms than they were with simplex forms, and in vocabulary recall. These results support previous research (Kempe & Brooks, 2001) that reported positive benefits for diminutive forms in the acquisition of Russian gender agreement among adult learners. In the present study, however, the benefits of the diminutive forms also appear to extend to case-marking, an area of grammar that, as noted, is notoriously different for L2 learners to acquire (Larsen-Freeman, 2010; White, 2009).

Given the difficulty of acquiring case markings, the relatively low levels of accuracy should not come as a surprise, particularly when compared with previous research, such as that of Grey et al. (2014), who reported similar findings. It is important to note, however, that the training phase of Grey et al (2014) consisted of a single exposure session that lasted approximately 30 minutes. By contrast, Brooks et al. (2011) had a comparatively extensive training phase of six one-hour training sessions. In addition, these training sessions involved a number of tasks, including consciousness-raising tasks, such as the provision of feedback on participants' performance (see Leow, 2015a, 2015b; Leow & Hama, 2013 for discussions). Despite this, performance on both case marking and gender agreement remained relatively low. The authors noted that

> this suggests that many adult learners were unable to discover the grammatical categories on their own, even though the inflectional paradigm was fully transparent (i.e., there were no irregular nouns)…thus, in the absence of explicit teaching, adult learners will often fail to notice grammatical patterns and dependencies on their own (Brooks et al., 2011, p. 1165).

This statement is predicated on previous research (Leow, 1997, 1998, 2000) that has demonstrated that noticing and higher levels of awareness are facilitative for language acquisition. In the case of Brooks et al. (2013), however, such a statement is speculative: In the absence of data, can it be said with any certainty that it is awareness, and not another variable, that accounts for variation amongst the participants' performances?

4.4.4. Brooks and Kempe (2013)

Brooks and Kempe (2013) reanalysed the data from Brooks et al. (2011), including the individual difference measures and retrospective verbal reports, to arrive at a better understanding of what might account for the variance in the degree to which participants were able to acquire the case marking and gender categories of the miniature Russian system. Several individual difference measures were considered, including an auditory sequence-learning task (Misyak & Christiansen, 2012), phonological short-term memory via a non-word repetition task (Gupta, 2003), verbal working memory capacity via a reading span task (Daneman & Carpenter, 1980), and nonverbal intelligence via a culture fair test (Cattell & Cattell, 1973). Further information about the foreign language backgrounds of the participants was elicited. Finally, retrospective verbal reports were taken and coded according to the amount of information participants were able to report. For example, participants received a score of zero if they were not able to provide any information. They received a score of one if they mentioned that the endings of the foreign words changed depending on whether the object was moving towards or away from another object, or if they listed two or more vowels that could occur at the end of words. Participants received a score of two if they were able to state some form of rule for the changing inflections, or if they were able to identify three or more of the inflections in the study. Thus, this research builds upon the findings of Brooks et al. (2011) described above, as well as on previous work by these researchers that has pointed towards a link between verbal working memory capacity and the acquisition of both L2 lexis and gender categories (Kempe et al., 2010).

The results of an initial regression analysis, which excluded metalinguistic awareness as a predictor variable, indicated that the learning of gender was best predicted by the participants' familiarity with a language that has a similar gender-marking system. Vocabulary learning was best predicted by the reading span task and the non-word repetition task, whereas the learning of the case marking system was predicted by the nonverbal intelligence task. Furthermore, the auditory sequence learning task predicted the level of metalinguistic awareness of the learners. A second regression analysis that included metalinguistic awareness as a possible predictor variable was conducted. The results of this analysis showed that the strongest predictor of the successful learning of case-marking was the extent to which learners became aware of the patterns in the input. In other words, metalinguistic awareness mediated the effects of both nonverbal intelligence and auditory sequence learning. Regarding vocabulary learning, however, the non-word repetition task remained the strongest predictor, whereas performance in the reading span task was no longer significant.

In summary, the results of Brooks and Kempe's (2013) study indicated that the degree to which participants were able to acquire an L2 case-marking system was strongly influenced by their ability to verbalise the rules governing the case-marking system. Similar to the results of Robinson (2002, 2005) and Grey et al., (2014), these results clearly show that explicit knowledge of inflectional morphology is linked to improved performance. Furthermore, these results suggest that awareness plays an important, if not vital, role in the learning of inflectional morphology, regardless of the intentionality of the learning condition. Taken together, these results could be interpreted in the light of the weak version of Schmidt's noticing hypothesis, in which awareness is facilitative for language acquisition, and perhaps even necessary for certain grammatical structures (Schmidt, 2010; see Section 2.3). Following this argument, the immediate question is whether awareness is necessary for the acquisition of L2 inflectional morphology and, furthermore, whether the difficulties learners face in learning inflectional morphology (Larsen-Freeman, 2010) could be at least partly attributed to issues of attention and awareness in SLA.

Despite the consistency of these results with those of other research in this area (Grey et al., 2014; Robinson, 2002, 2005) in pointing towards an important, if not necessary, role for awareness in the learning of L2 case markings, there are a number of limitations to this study. Most notably, as in previous research in this area, Brooks and Kempe (2013) utilised retrospective verbal reports as the sole measure of awareness. Although the use of verbal reports in this study did allow for a finer-grained interpretation of the results, issues related to the veridicality of these results persist. In particular, the results of retrospective verbal reports provide little information on their own about the degree to which participants developed implicit and explicit knowledge during the training phase and utilised this knowledge in the testing phase.

4.5. Summary

In summary, despite the attention paid in recent years to both the incidental learning of L2 grammar and the acquisition of L2 morphology, few studies have directly examined the acquisition of L2 case marking under incidental learning conditions after a minimal amount of exposure. The results of three studies that have set out to address this gap (Grey et al., 2014; Robinson, 2002, 2005) paint an unclear picture concerning the degree to which L2 case markings can be acquired following incidental exposure. In Robinson's (2002, 2005) studies, a significant learning effect was found for only one of three morphological case markers. In Grey et al. (2014), evidence of learning was only obtained from delayed, but not from immediate, post-tests. Furthermore, in these studies, verbal reports provided little evidence to substantiate the claim that L2 case markers can be acquired in the absence of conscious awareness. The findings of these studies are corroborated by an additional investigation by Brooks and Kempe (2013), who found that verbalisable, explicit knowledge was the strongest predicator for learning an L2 case-marking system, albeit under incidental learning conditions.

Building and expanding on the work of Grey et al. (2014, 2015), Robinson (2002, 2005), and Brooks and Kempe (2013), this thesis set out to investigate further the extent to which incidental learning conditions can promote the acquisition of L2 case marking. In addition, these experiments explored whether knowledge acquired after incidental exposure is implicit or explicit in nature. It should be reiterated here that previous research by Grey et al. (2014) did include subjective measures of awareness for tasks measuring knowledge of the L2 word-order system, but not for the picture-matching task, which was designed to measure knowledge of case marking. Therefore, a methodological improvement of this research was that, unlike Grey et al., Robinson, and Brooks and Kempe, the experiments described here utilised subjective measures of awareness in addition to retrospective verbal reports (see Rebuschat, 2013 for a review). This allowed for a fuller understanding of the type of knowledge that results from incidental exposure. Thus, this research not only contributes to the discussion surrounding the issue of incidental acquisition of L2 morphology, but also informs the thorny debate of whether a dissociation between learning and awareness is possible. In particular, this study examines whether it is possible to learn a specific feature of L2 grammar, namely case marking, in the absence of conscious awareness of that which has been acquired.

On a methodological level, this thesis presents a number of innovations. Firstly, Pilot Studies 1 and 2 and Experiments 1 and 2 have incorporated subjective measures of awareness within an AGL paradigm to investigate the incidental learning of L2 case markings. As noted, such an addition represents a methodological improvement to previous research in this area because it allows for the triangulation of data from multiple measures of awareness. Secondly, Experiment 3 utilises both direct and indirect measures of learning to further triangulate data regarding the conscious status of acquired knowledge. The indirect measure of learning is presented in the form a self-paced reading task, a novel methodology within this area of enquiry. The direct measure is a multiple-choice sentence completion task accompanied by subjective measures of awareness. Taken together, these methodological innovations allow for a more refined, multi-faceted perspective on the conscious status of the knowledge acquired as a result of incidental exposure. Such a perspective will not only serve to inform future research in this area, but will also add to the ongoing debate on the role of implicit and explicit processes in SLA.

4.6. Research Questions

The following research questions were formulated.

RQ 1: To what extent can L2 case marking be acquired under incidental learning conditions?

RQ 1a: How is this knowledge reflected in direct measures of learning, such as grammaticality judgement tests and multiple-choice tasks?

RQ 1b: How is this knowledge reflected in indirect measures of learning, such as a self-paced reading task?

RQ 2: What type of knowledge is acquired as a result of this exposure; implicit or explicit?

RQ 2a: How is/are implicit and/or explicit knowledge reflected in direct and indirect measures of learning?

RQ 2b: How is/are implicit and/or explicit knowledge reflected in retrospective verbal reports?

RQ 2c: How is/are implicit and/or explicit knowledge reflected in subjective measures of awareness?

CHAPTER 5

PILOT STUDIES 1 AND 2

This chapter reports on two pilot studies that aimed to test and validate the experimental materials and procedure to be used as part of the experiments in this thesis. Both Pilot Studies 1 and 2 utilised a semi-artificial language within the AGL research paradigm to examine the incidental learning of an artificial morphological system. Although semi-artificial languages have grown in popularity in recent years within SLA, their use is not without methodological challenges. One important but rarely discussed aspect of this line of research is the important step of establishing an initial learning effect. That is, in order to work with artificial systems, researchers need first to ensure that the systems are actually learnable by participants under experimental conditions. Once the learning effect has been observed, it is then possible to observe, in subsequent experiments using the same artificial language, how different manipulations impact on learning (see Rogers, Révész, & Rebuschat, 2015 for a brief discussion). As such, one of the primary aims of the pilot studies was to establish an initial learning effect.

This chapter is organised as follows. First, a brief overview of the purpose and experimental design of Pilot Study 1 is given. Next, a more detailed description of the population, methods, materials, and experimental procedure is provided, with the aim of providing sufficient detail so that future replication is possible. Before discussing the results of this pilot study, a brief rationale is given for the various statistical analyses utilised. This is followed by an interim discussion, in which the effectiveness of this experimental design is evaluated critically. This general outline is then repeated for Pilot Study 2. This chapter concludes with a brief summary of both pilot studies, which includes a discussion of the implications of these studies for the experimental design of *the experiments in this thesis.*

5.1. Pilot Study 1

Pilot Study 1 focused on the incidental learning of morphological rules, specifically the nominative marker *–a*, the accusative marker *–u*, and the instrument marker *–ou,* of a semi-artificial language based on Czech morphology. Czech morphology was selected as the target for the following reasons. First, Czech is a more morphologically rich than English, and it contains a rich case-marking system that is not present in the L1 of the participants of this study. Secondly, it is a language that would be unfamiliar to the majority of the members of the population pool. As such, using an unfamiliar language such as Czech allows for controlling for prior knowledge on the part of the participants.

5.1.1. Overview of Czech Language

Languages vary with regard to their use of morphology and their morphological structures[16]. Within the field of linguistic typology, languages have been classified historically based on the amount of accumulated information that words within a language are able to hold. Analytic languages, such as Mandarin, have a low morpheme to word ratio and rely on word order to establish grammatical relations within a sentence (Clark, 2001). In contrast, languages in which words are typically comprised of more than one morpheme are referred to as synthetic

[16] For overviews of linguistic typology, see Comrie (2009), Song, (2011) and Velupillai, (2012).

languages. These languages often utilize affixes to establish morphosyntactic relationships (Villupillai, 2012). Extremely synthetic languages, such as Turkish, are referred to as polysynthetic languages. Within these languages, words are comprised of a complex array of morphemes that can often constitute entire clauses.

It is important to stress that the analytic to synthetic to polysynthetic distinctions are relative rather than absolute in that a language never really belongs to one category or another. (Brown, 2011; see also Sapir, 1912). The English language, for instance, is more of an analytic language, despite the presence of synthetic elements (Slobin, 1997; Vellupillai, 2012). This is because it strongly relies on word order and the use of auxillary words, rather than synthesis and inflectional morphology, to establish grammatical relations within a sentence. In contrast, Slavonic languages, such as Russian, Polish, and Czech are highly synthetic in that they contain complex and fully-developed case-marking and gender-agreement systems (Comrie, 2011; Lukavský & Smolík, 2009). The following section will provide a brief overview of Czech, a highly synthetic Slavonic language that is the basis of the target linguistic structure of the experiments throughout this thesis.

<u>Overview of the linguistic features of Czech</u>

As noted above, Czech is a highly synthetic language that belongs to the western group of Slavonic languages (de Bray, 1980; Hanson, 2011). Morphologically, Czech is similar to other syntactic languages in that it contanis a rich, morphologically marked word class system in which verbs, nouns, adjectives, personal pronouns and numerals all carry inflectional markers (Hanson, 2011). Like most Slavonic languages, Czech has six syntactic cases as well as a vocative form, (Hanson, 2011; see also description of Russian in Slobin, 1997). This case system is used to mark grammatical relationships within a sentence (Short, 2009). As part of this sytem, Czech includes a number of case-marking paradigms. Which pattern a particular noun follows depends on a number of different variables, such as the gender of the noun in question (masculine, feminine, or neutral) and whether the ultimate consonant in the stem is hard or soft (Comrie, 2011; de Bray, 1980). Declension patterns further vary for masculine nouns depending on whether the nouns represent animate or inanimate objects (Janda & Townsend, 2000; Short, 2009). An example of a Czech declension pattern can be seen in Table 1.1 below.

Table 1.1. *Example of Declension of Feminine Hard Nouns in Czech*

	Czech Singular	Czech Plural	English Equivalent	English Example
Nominative	žena (woman)	Ženy (women)	Subject	The <u>woman</u> is here.
Vocative	ženo	ženo	To call out	Hey! <u>Woman</u>!
Accusative	ženu	ženy	Object	He kissed the <u>woman</u>.
Genitive	ženy	žen	Direction / without	From/without the <u>woman</u>.
Dative	ženě	ženám	Indirect Object	He gave the <u>woman</u> the gift.
Instrumental	ženou	ženami	with	He went with the <u>woman</u>.
Locative	ženě	ženách	Direction / Location	He walked towards the <u>woman</u>.

Adapted from Short (2009); de Bray (1980), & Petr (1986).

The basic word order of Czech is SVO (i.e., subject + verb + object). However, although SVO is standard, Czech syntax is not limited to an SVO pattern. For example, an OVS pattern is also grammatically possible within the Czech language (Lukavský & Smolík, 2009; Petr, 1986; Short, 2009). One reason for this flexible word order is because syntactic relationships are fully realised via the morphological case markings in the Czech language (see Table 1.1 above). As such, a speaker of Czech would be able to identify the subject and object of a sentence based on the morphological marker that these words carry. This means that "A cat killed a mouse" and "A mouse was killed by a cat" can be realized as "Kočka zabila myšku" and "Myšku zabila kočka", respectively. In this example, the inflection –a indicates that kočka is the subject of the sentence and the inflection –u indicates that myšku is the object, regardless of their relative positions within the sentence. As can be seen in these examples, sentence functions are clearly realized by the case markers. This allows for a freer word order than can be found in more analytic languages, such as English (Hanson, 2011; Short, 2009). However, it is important to note that free word order does not equate to a totally "random" word order. Much research has been carried out on the linguistic functions of word order in Czech. Furthermore, it should be noted that Czech has quite strict word order regarding some grammatical features not relevant to the present thesis, such as enclitics (see, e.g., Short, 2009).

5.1.2. Methods

This experiment followed the AGL paradigm to address the following research questions:

1). To what degree can L2 case markings be learned under incidental learning conditions?

2). What type of knowledge is acquired as a result of this exposure; implicit or explicit?

As noted in Section 3.1 above, the AGL paradigm typically entails three stages: a training phase, a testing phase, and a debriefing session. In the training phase, participants are exposed to stimuli under incidental learning conditions. This training phase is followed by a surprise testing phase, of which participants are not informed in advance. Following the testing phase, the debriefing session consists of either a written or an oral interview in which the participants are asked a series of questions to determine their level of awareness of the target rule system.

5.1.2.1. Participants

Fifty-two native speakers of English with no background in any Slavic (or other morphologically rich) languages took part in Pilot Study 1. The participants were randomly assigned to an experimental group (n = 28; 18 female, 10 male) and a control group (n = 24, 15 female, 9 male). The participants were recruited via email invitations that were sent with the assistance of the programme administrators of several academic departments at a university in the United Kingdom. All participants were undergraduate students, but none were majoring in linguistics or foreign languages. The ages of the participants ranged from 18 to 24 years (M = 19.7, SD = 1.8). The data set for one participant was discarded due to a disruption in the experimental environment; another set was lost due to experimenter error. Each participant was offered £10 for taking part in this study.

5.1.2.2. Research Ethics

Prior to the outset of the data collection, this experiment was approved as exempt by the research support office of Lancaster University[17]. Informed consent to take part in the experiment was sought from each individual participant prior to the start of the training phase of this study. In the first instance, each participant read a

[17] This PhD project began at Lancaster University. As a result, ethical approval for the initial experiments in this thesis was granted by the research support office of Lancaster University. Experiments that took place after my transfer to The UCL Institute of Education, University College London, sought ethical approval from the Institute of Education's research support office.

description of the research project. Furthermore, each participant was given the opportunity to ask any questions or voice any concerns about the experiment or experimental procedure before giving his or her informed consent in writing. The experimenter stressed that participation in this study was voluntary, that participants could withdraw from the experiment at any time, without recourse of any kind (which included the loss of their financial reward), and without the need to give a reason for doing so.

It is important to note here that the true purpose of this study was not fully disclosed to the participants until the end of the experiment. This is because this study investigated the degree to which L2 case markings can be acquired under incidental learning conditions. If the participants had been made aware that this was the true focus of the study at the outset, this would have jeopardised the construct validity of the research design, in that the participants would have been more likely to pay attention to, and deliberately attempt to learn, the target structures. Therefore, in order to disguise the true purpose of this study, Pilot Study 1 was presented to the participants as a study that was investigating the learning of foreign language vocabulary (see Appendix A for the information sheet and consent form). However, the true purpose of the experiment was fully disclosed at the end of the experiment, and participants were given a further opportunity to ask questions and to withdraw their consent.

5.1.2.3. Stimulus Material

Training Set

A semi-artificial morphological system, based on Czech case marking, was used to generate the stimulus material for this experiment. As shown in Table 5.1, the system consisted of English phrases and a Czech noun, which was inflected according to one of three cases (nominative *–a*, accusative *–u*, instrumental *–ou*), depending on its function in the sentence (subject, object, instrumental).

Table 5.1.
Descriptions and Examples of the Three Morphological Categories in Pilot Study 1

Morphological Category	Syntactic category in English	Example
Nominative	Subject	*The britva cut David's face at the sink last night.*
Accusative	Direct Object	*Peter used a britvu in the bathroom today.*
Instrumental	Adverbial (meaning "to do/make something with an X")	*Anne cut her leg with a britvou in the morning.*

A total of 48 Czech nouns, all of which were foreign to the participants, were used in the training set. All the Czech words were regular, feminine nouns that end with the inflection *–a* in their nominative form. All the nouns followed the same pattern of declension. Only nouns with relatively "concrete" meanings were chosen in order to ensure that the nouns could be easily represented visually through images in the training phase.

For the training phase, 96 clip-art images were collected. Forty-eight of these images corresponded to the Czech words used in the training set, and 48 images were distractor images that did not correspond to any of the foreign words. Each of the 96 images was used three times in the training phase of the study. The distribution of the distractor images was balanced throughout the training phase so that they did not occur more than once with any particular Czech noun. The complete set of images used in this thesis can be found in Appendix J.

Three stimulus sentences were created for each of the 48 Czech nouns in the training set. Of these three sentences, one sentence was written in such a way that the Czech noun occurred in the nominative case, one sentence included the noun in the accusative case, and one sentence had the noun marked for the instrumental case. In summary, each Czech noun occurred three times in the training set, each time in a different sentence and each time with a different function and case marking. This resulted in a total of 144 sentences–48 sentences for each of the three morphological categories. Please see Appendix C for a complete list of the stimulus array used in this experiment.

In addition to the inflected Czech noun, the word order in the sentences was scrambled according to four syntactic patterns (see Table 5.2, below, for templates of syntactic patterns and example sentences). There were a number of reasons for scrambling the word order in these sentences. Firstly, scrambled word order is more reflective of natural languages with rich morphologies. As in Czech, such languages rely on case endings rather than on word order to mark functions within the sentence. Secondly, the syntactic patterns allowed for controlling the position of the foreign or Czech word within the sentence. This ensured that the position of the foreign word could not serve as a reliable indicator of its function within the sentence. Finally, it has been argued that stimuli need to be sufficiently complex for the development of implicit knowledge within short training contexts (Reber, 1993; Rebuschat, 2013). The reasoning behind this claim is that if the stimuli are relatively simple, it will be easy for participants to ‘crack the code’ and to develop a conscious understanding of the rules governing the stimulus domain. Furthermore, if the stimulus domain appears simple, it could potentially encourage the use of explicit learning strategies, such as rule-search behaviour, thus jeopardising the construct validity of the incidental learning conditions. In summary, given the relatively limited scope of the morphological system utilised in the present study, varying the syntactic patterns arguably increases both the ecological validity of the stimulus domain and the internal validity of the research design.

Table 5.2.
Template, Sample Sentences, and Frequencies for the Four Syntactic Patterns in Pilot Study 1

Pattern	*Template*	*Frequency*
Pattern 1	$[[AP]_{TEMP} > [NP]_{OBJ} > [VP] > [NP]_{SUBJ} > [PP]]$	
Nominative	*Last summer the grass ate the koza in the field.*	(12)
Accusative	*Last month the kasu opened Patrick with the key.*	(12)
Instrumental	*Some time ago John scared the child with a zrudou.*	(12)
Pattern 2	$[[AP]_{TEMP} > [NP]_{SUBJ} > [VP] > [NP]_{OBJ} > [PP]]$	
Nominative	*Last year the prodejna shipped goods to the shoppers.*	(12)
Accusative	*All week the builder took his vrtacku to work.*	(12)
Instrumental	*Today the wooden board cut he with a pilkou.*	(12)
Pattern 3	$[[NP]_{OBJ} >[VP] > [PP] >[AP]_{TEMP} > [NP]_{SUBJ}]$	
Nominative	*David's face cut at the sink at night the britva.*	(12)
Accusative	*The zahradu planted with fruits weeks ago Beth.*	(12)
Instrumental	*The dishes washed with a myckou last night Peter.*	(12)
Pattern 4	$[[NP]_{SUBJ} >[VP] > [PP] >[AP]_{TEMP} > [NP]_{OBJ}]$	
Nominative	*The kocka killed with its teeth this morning the bird.*	(12)
Accusative	*The cat chased in the house in summer the mysku.*	(12)
Instrumental	*Sarah shot with a flintou weeks ago a bird.*	(12)

All the sentences in the training phase were written so that an approximate meaning of the foreign word could be inferred by the participants from the rest of the sentence. In the construction of the training and testing sets, care was taken to control for the length of the sentences. Each sentence had exactly 12 syllables, and a frequency analysis confirmed that the number of words per sentence was not a reliable indicator of morphological category during the training phase, F (2, 141) = 1.322, p = .27, η^2 = .02.

In addition, all English words in the stimulus materials were among the 2000 most frequent English words, as determined by Lextutor's vocabulary profile programme (Cobb, 2012). The sentences were also the same in terms of lexical diversity (types per tokens, $M = .04$, $SD = .01$), lexical density (content words per total number of words, $M = .59$, $SD = .01$), and average word length (number of syllables per word, $M = 1.30$, $SD = .04$).

Testing Set

The testing set of this experiment consisted of 48 new sentences. Half of these sentences (24) were transfer items; in other words, they consisted of novel Czech words and sentences that had not occurred in the training set. The other half of the sentences (24) were partially trained items in that the Czech word in these sentences had occurred in the training set, but in a different sentence context.

All items in the testing set were designed with the same considerations as were those in the training set, controlling for the total number of syllables per sentence (12), as well as for lexical and syntactic complexity. A frequency analysis of the testing set indicated that the average stimulus length was the same for the grammatical and ungrammatical items (9.04 words per sentence). There was also no significant difference between the sentence length of the stimulus materials used in the training phase ($M = 9.15$ words) and in the testing phase ($M = 9.04$ words), t (66.901) = 1.068, $p = .29$, $d = .18$. This indicates that sentence length could not serve as a reliable predictor of grammaticality during the testing phase.

The ungrammatical items in the testing set were generated by replacing the correct case marking with one of the other two case markers that had also been present in the training set. Incorrect case markings were balanced across the testing phase. Of the 48 total items in the testing phase, 16 were nominative (-*a*), 16 were accusative (-*u*), and 16 were instrumental (-*ou*). Eight of the nominative items were grammatical and eight were ungrammatical. Of the eight ungrammatical nominative case-items, four were created by replacing the nominative marker (-*a*) with the accusative marker (-*u*), and four by replacing the nominative marker with the instrumental marker (-*ou*). The same procedure was followed to create the accusative and instrumental items. In summary, care was taken to ensure that participants could only make correct judgements in the testing phase if they were able to identify instances of correct and incorrect case marking. Examples of the sentences used in the testing set can be found in Appendix D.

5.1.2.4. Procedure

As noted above, the experiment consisted of three phases: a training phase, a testing phase and a debriefing session, during which the participants provided retrospective verbal reports. The experimental group completed all three of these phases; the control group took part in the testing phase and the debriefing session. The experiment took place in a private office on campus. The training and testing phases of the experiment were delivered via a Dell Thinkpad T410 laptop computer, using the stimulus presentation software Superlab 4.5 (Cedrus Corp, San Pedro, CA). All audio for the experiment was played through a set of Audio-Technica headphones with noise-cancellation functionality. Following the testing phase, participants completed a short debriefing questionnaire, followed by an oral interview.

Training phase

In the training phase, participants were exposed to the stimulus material under incidental learning conditions; in other words, they did not know they were going to be tested. Furthermore, following common practice in recent studies on incidental learning (Hamrick, 2013; Rebuschat & Williams, 2012), the training phase was deliberately designed to disguise the real purpose of the training task. Participants were told that they were going to take part in a study on learning foreign language vocabulary. Their task was to listen to a sentence, then match the meaning of the foreign (Czech) word (such as *žehličkou*) to one of two pictures (see the description of the images, above) displayed on the monitor (for example, an iron or a broom). Participants were given no feedback on the accuracy of their decisions. Figure 5A, below, illustrates the training procedure.

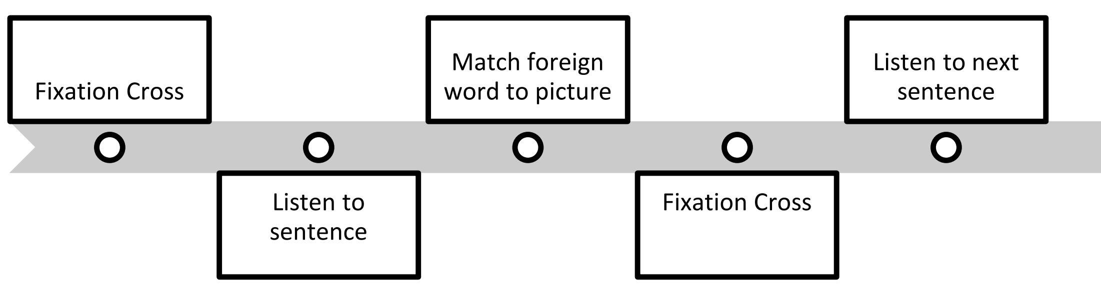


Figure 5A. *Training Procedure in Pilot Study 1.*

At no point during the training phase were participants informed that the foreign nouns were inflected for case, nor that they would be tested afterwards. Participants listened to all 144 sentences without a break. These sentences were presented in a different, randomised order for each participant. The entire training phase took about 25 minutes to complete. Please see Figure 5B, below, for examples of the slides used in the training phase of Pilot Study 1.

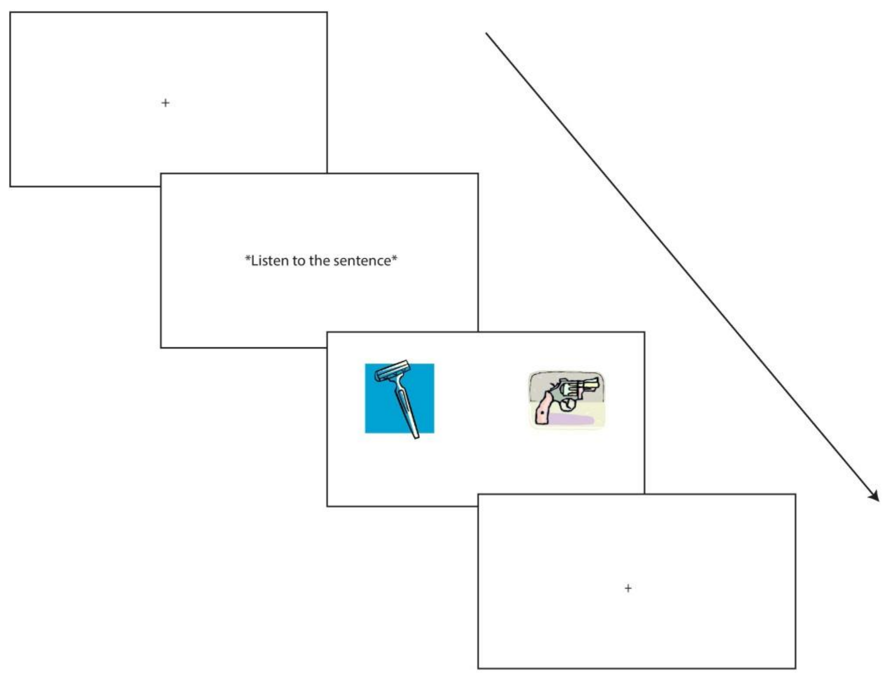


Figure 5B. *Example of Slides Used During the Training Phase of Pilot Study 1*

Testing Phase

The testing phase for this experiment consisted of a 48-item grammaticality judgement task. Following the training phase, participants in the experimental group were informed that the sentences in the previous section were not arbitrary, but were part of a complex system. They were then told that they would listen to 48 new sentences, half of which belonged to the same system, and half of which did not. For each test sentence, participants had to decide if the sentence belonged to the same system as quickly as possible. No feedback was provided on the accuracy of the participants' decisions. See Figure 5C below for an illustration of the testing procedure, and Figure 5D, below, for the slides that were used in Superlab for the testing phase of the experiment.

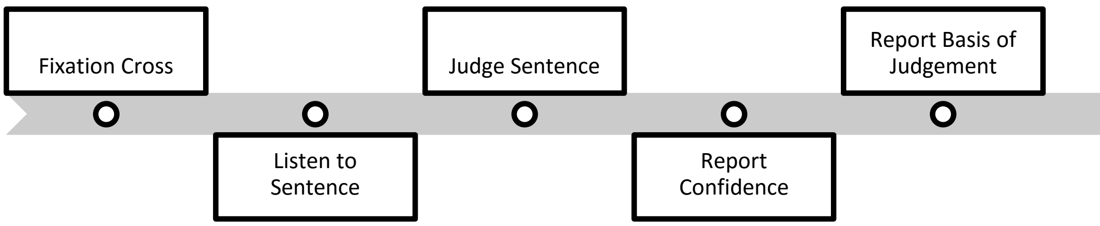


Figure 5C. *Testing procedure in Pilot Studies 1 and 2, and Experiments 1 and 2.*

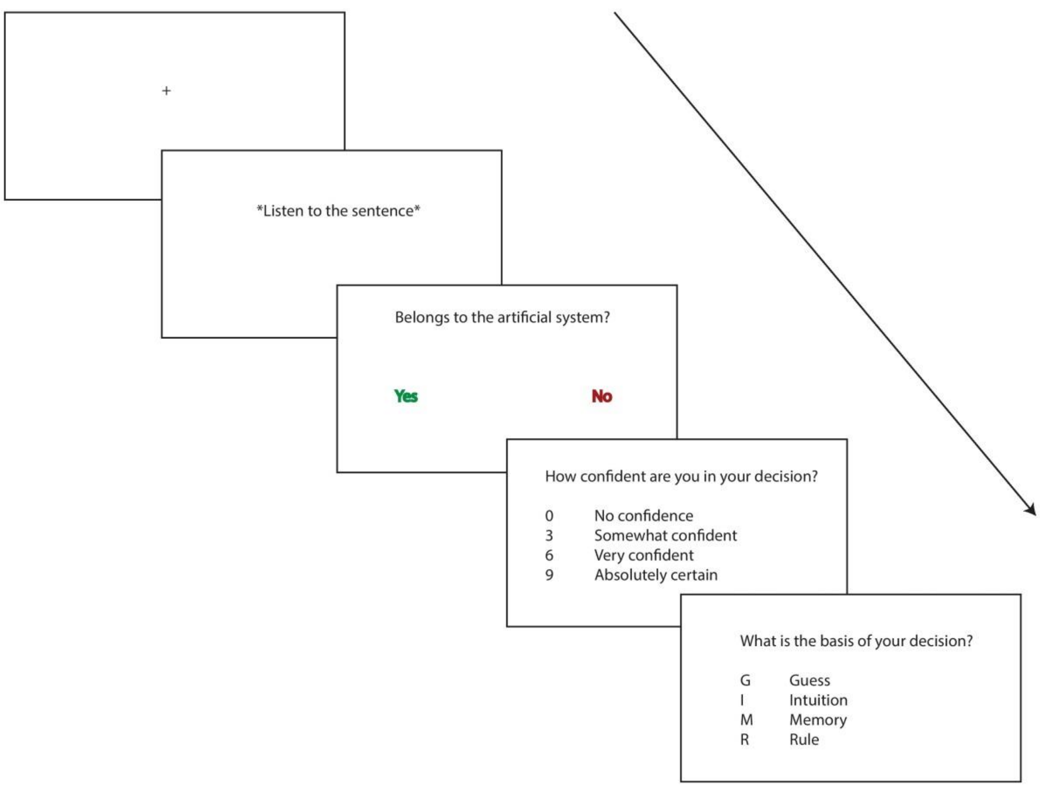


Figure 5D. *Example of Slides Used in the Testing Phase of Pilot Study 1*

At the end of the experiment, the participants were also prompted to describe, in both a written questionnaire and as part of a follow-up oral interview, any rules or patterns they might have noticed. In the oral interviews, the researcher asked the candidates to elaborate on their responses in the written questionnaire. Finally, the researcher explained the underlying rule system and asked the participants again if they had figured it out or had any intuition about it at any point during the experiment.

5.1.2.5. Statistical Analyses

Performance on the grammaticality judgement task was initially analysed using mean accuracy rates and logit mixed-effects models. As mixed-effects models will be unfamiliar to many SLA researchers, some background to these statistical procedures is provided in the section below. Before examining the subjective measures of awareness to determine the degree to which the exposure phase resulted in implicit and/or explicit knowledge on the part of the participants, it must first be established that the incidental training conditions produced a significant learning effect. This is necessary because subjective measures of awareness are designed to determine whether knowledge acquired during an experiment is implicit or explicit in nature. If the results of the experiment do not indicate a significant learning effect, it can be assumed that no knowledge, implicit or explicit, was acquired as a result of incidental exposure (see Rebuschat, 2013).

5.1.2.6. Mixed-effects Statistical Models

Within the psycholinguistic literature, there has been a surge of interest in mixed-effects statistical models to the degree that mixed-effects models are now considered the preferred type of analysis when dealing with judgement data, such as on a GJT (Schütze & Sprouse, 2013; see also Ionin & Zyzik, 2014 for a review of judgement tasks in SLA). This interest in mixed-effects modelling followed a special issue of the *Journal of Memory and Language*, which contained several articles espousing the use of this type of analysis (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Quené & van der Bergh, 2008). As argued within these articles, there are a number of advantages in the use of mixed-effects models in psycholinguistic research (see Linck & Cunnings, 2015 for a discussion). Of particular relevance to the present thesis is that these models address what is referred to as the "language-as-fixed-effect fallacy" (Baayen et al., 2008; Clark, 1973; Coleman, 1964). This argument holds that, although significance testing is typically carried out in order to determine if the results of a study can be generalised beyond the participants who took part in it, little is known regarding whether the results of experiments can be generalised beyond the linguistic items included in them. In other words, "there is little statistical evidence that such studies could be successfully replicated if a different sample of language materials were used" (Coleman, 1964, p. 219). Mixed-effects models are able to treat both subjects and items as random variables, thus allowing for a better generalisation of the results (Gagné & Spalding, 2009).

At this point, it is worth defining some terms that will aid in understanding mixed-effects statistical models. SLA researchers will be familiar with the terms *dependent variable* and *independent variable*. *Dependent variables* are the variables of focus in a particular experiment. *Independent variables* are variables that are selected and manipulated by the researcher to determine the effect these manipulations have on the dependent variable (Brown, 1998). Mixed-effects models, on the other hand, adopt different terminology and are discussed in terms of *fixed effects* and *random effects* (see Cunnings, 2012, for an overview of mixed-effects models in second language research). In mixed-effects models, fixed effects are typically the points of interest in a study, and are often equivalent to the independent variable in an experiment. Random effects are variables originating from the random selection of subjects (such as age) and/or items (such as word length) included in a particular experiment (Quené & van den Bergh, 2008).

Within this thesis, mixed-effects models were implemented using R (R Core Team, 2014) and lme4 (Bates, Maechler, Bolker, & Walker, 2014). All models were set up using the maximal structure advocated by Barr, Levy, Scheepers and Tily (2013) in which the statistical model always includes a corresponding random slope for each fixed effect. Furthermore, these models largely followed the procedures laid out by Cunnings (2012; see also Baayen et al., 2008; Jaeger, 2008; Linck & Cunnings, 2015), and which have been adopted by a number of recent studies in SLA (de Zeeuw, Verhoeven, & Schreuder, 2012; Jackson, 2014; Linck, Kroll, & Sunderman, 2009; Sonbul & Schmitt, 2013). Following Cunnings and Sturt (2014) and Barr et al. (2013); if the maximal model failed to converge, the random effect that accounted for the least variance in the data was removed until the model converged.

There were several steps in the process of analysing the data using these mixed-effects models. First, a series of analyses were carried out to determine which fixed effects showed a significant relationship with the dependent variable. Within standard regression models, such a procedure is carried out to avoid multicollinearity, or the correlational between predictor variables in the model. A benefit of avoiding multicollinearity is that a reduction in the number of factors entered into the model raises the statistical power of a study; in other words, the probability of detecting statistical significance (Brown, 2015).

Although mixed-effects models are considered robust against violations of assumptions that constrain traditional regression analyses, such as normality of distributions, homoscedascity and sphericity (Quené & van den Burgh, 2008), it is important to ensure that collinearity between the fixed effects is not impacting on the results of the study (Baayen, 2008). Within standard regression models, such an analysis is typically carried out by producing a correlational matrix, for example by examining the bivariate correlations within a regression model (Jeun, 2015; Pallant, 2013; see also Brooks & Kempe, 2013 for an example). Highly correlated predictor variables are then either removed from the model, or are combined to form a single predictor variable (Jeun, 2015).

One problem with such an approach is that a typical correlational analysis does not control for subject and item effects in the same manner as do mixed-effects models. Therefore, an alternative approach in which the predictor variables were analysed against a null model using the $\chi 2$ statistic of a likelihood ratio test was adopted here. Significant predictor variables were then compiled into a final model, which was formally examined for multicollinearity using Cohen's kappa (Cohen, 1968). Such a methodology allows for an examination of the relationship between the dependent variable and predictor variables, albeit on an individual basis, while controlling for subject and item effects in the data.

This procedure was carried out as follows. First, a null model was built, which analysed the dependent variable in terms of crossed random intercepts of subjects and items. This model was then compared with a second model that included a fixed effect as well as a corresponding random slope using the $\chi 2$ statistic, as stated above. If this ratio test indicated a significantly improved fit for the second model, this result was then interpreted as a significant relationship with the dependent variable.[18] This was done on a one-by-one basis with each of the fixed effects of interest.

Next, the fixed effects that did show a significant relationship with the dependent variable were all entered into a new model, which include random slopes for each of these fixed effects. This model was then examined in order to determine if the "complete" model warranted the inclusion of all of the fixed effects. This was done by removing the fixed effects one at a time, and then comparing this model with the complete model. If a likelihood ratio test indicated that the exclusion of a particular fixed effect led to an improved model fit, the fixed effect in question was then removed from further analyses. This procedure was carried out on an individual basis for each fixed effect in order to arrive at the best fitting model. In all cases, the best fitting model was formally analysed for multicollinearity by computing Cohen's kappa (Cohen, 1968). Following Baayen (2008; see also Belsley, Kuh, & Welsch, 1980) a kappa value of less than 6 is indicative of low collinearity, between 6 and 30 of moderate collinearity, and over 30 of severe collinearity.

[18] It is important to note that $\chi 2$ can also be used to interpret whether a fixed effect is significant (see Barr et al., 2012). However, as it has been argued that this method is anti-conservative (Pinheiro & Bates, 2000), particularly with smaller sample sizes (Barr et al., 2012). Here, it is interpreted as a significant relationship between the predictor variable and the dependent variable.

The next step in this process was to analyse the summary statistics for the best-fit model. Following Linck and Cunnings (2015; see also Baayen et al., 2008), significance here is interpreted as absolute *t*-value or *z*-value ≥ 2.0[19]. It is also important to discuss briefly how significant main effects, as indicated by the absolute *t*-value, should be interpreted. If a fixed effect reaches significance within the model, then this indicates that the effect in question is significant in accounting for variance in the overall dataset. As an example, if a model indicates that the fixed effect of group is significant, this points towards a significant difference in accuracy between the two variables within two categories within the fixed effect of group (experimental and control). In addition to significant main effects, the mixed-effects model can also indicate a significant interaction between two fixed effects. It is important to note again that a significant interaction does not indicate a correlation between fixed effects, but that the effects interact to explain variances in the data. For example, if the model results in a significant interaction between group and case, then this result points towards differences between the experimental and control groups' performances with regard to the different case markers assessed in the testing phase. In the event of significant interactions, post-hoc Tukey tests were conducted to ascertain the exact nature of the relationship. These tests were carried out using the GLHT command of the R package "multcomp" (Hothorn, Bretz, & Westfall, 2008).

Following the example provided by Baayen (2008), effect sizes for the logit-mixed effects model were calculated using the C index of concordance and the Hmisc

[19] For logit mixed-effects analyses, the lmer package automatically approximates *p*-values based on the *t*-value or *z*-value. However, this is not provided for linear mixed-effects models (as used in Experiment 3).

package (Harrell & Dupont, 2015). C index values range from .5 to 1, with .5 representing no fit and 1 representing a perfect fit of the data. A C-index of .7 is interpreted as a moderate fit, .8 and above as a good fit, and .9 and above as an excellent fit for the data (Gries, 2013). Effect sizes for the post-hoc Tukey tests were calculated using Cohen's *d*. Following Plonsky and Oswald's (2014) recommendations, Cohen's *d* of .40 is considered small, .70 medium, and of 1.00 and higher as a large effect size.

5.1.3. Results

5.1.3.1. Retrospective verbal reports

An analysis of the retrospective verbal reports (both written and oral) revealed several important points. Firstly, with regard to the stimulus set, all the participants (28/28) commented on the syntactic patterns, with a number of comments related to the fact that some of the patterns were “strange” or “different from English.” Several of the participants (12/28) suggested that these sentences were more challenging to understand. In other words, they noted that the varying syntactic patterns made it more difficult to comprehend the overall meaning of the sentence. With regard to the case markers in particular, none of the participants (0/28) in the experimental group were able to verbalise the target morphological rules at the end of the experiment, even when the experimenter prompted them to guess. In addition, after the experimenter explained the rules to the participants and asked if they had thought of these rules previously, none of the participants stated that they had done so. Also, only around half (13/28) of the participants reported noticing the endings of the foreign words and the fact that the endings were different on different words. Furthermore, when prompted to guess why the endings of the foreign words were changing, slightly more than half of the participants (16/28) guessed that the changing inflections were related to noun gender in some manner.

5.1.3.2. Picture-matching task

The analysis of the participants' performance on the picture-matching task in the training phase of the experiment indicated that they were able to match the foreign word with its corresponding picture correctly and with great accuracy (*M* = 95.93%, *SD* = 3.50%).

5.1.3.3. GJT

Descriptive statistics were compiled to examine the performance of the experimental and control groups on the GJT. These statistics were generated both with regard to overall performance, and with regard to performance in the various sub-categories of the GJT, as can be seen in Table 5.3 below.

Table 5.3.
Results of Pilot Study 1: Descriptive Statistics (%) for Experimental and Control Groups across Sub-components of GJT

Item Type	*Descriptive Statistics (%)*		
Group	M	SD	SE
Grammatical			
Experimental	50.80	10.54	2.07
Control	47.47	10.92	2.23
Ungrammatical			
Experimental	55.45	8.56	1.68
Control	53.65	10.38	2.19
Nominative			
Experimental	49.76	11.25	2.21
Control	50.00	11.06	2.26
Accusative			
Experimental	53.85	8.12	1.59
Control	51.56	13.82	2.82
Instrumental			
Experimental	55.96	10.85	2.13
Control	51.75	13.83	2.82
Trained Items			
Experimental	50.80	7.17	1.41
Control	48.78	8.99	1.83
Transfer Items			
Experimental	55.45	7.70	1.53
Control	52.43	7.05	1.44
Pattern 1			
Experimental	56.09	9.88	1.94
Control	52.79	16.97	3.46
Pattern 2			
Experimental	51.92	9.22	1.81
Control	52.78	9.41	1.92
Pattern 3			
Experimental	54.49	11.84	2.32
Control	46.18	14.94	3.05
Pattern 4			
Experimental	51.60	13.75	3.00
Control	51.39	12.45	2.54

At a cursory glance, the overall performance of the experimental group (*M* = 53.13%, *SD* = 3.79%) appeared to be only slightly higher than was the performance of the control group (*M* = 50.60%, *SD* = 5.72%). However, examination of the sub-analyses revealed larger differences in the performance of the experimental versus that of the control group for Type 1 and Type 3 syntactic patterns, as well as for instrumental case items. In order to determine whether these differences in performance were statistically significant, a logit mixed-effects regression model (Jaeger, 2008) was carried out using the lme package (Bates et al., 2014) within R (R Core Team, 2014) to examine the relationship between group (experimental versus control) and accuracy in the GJT. At this point, initial analyses were carried out to determine which fixed effects had a significant relationship with the dependent variable; in other words, performance in the GJT via a series of logit mixed-effects analyses. For each of these analyses, a null model was constructed, which analysed performance on the GJT as a binary outcome (correct versus incorrect) in terms of subject and item as crossed random intercepts. Additional models were then built for each fixed effect. The fixed effects included in this model incorporated group (experimental versus control), grammaticality (grammatical versus ungrammatical test items), case (nominative versus accusative versus instrumental), syntax (Type 1, Type 2, Type 3, Type 4), and transfer (trained versus transfer items). Each of these models included the variable as a fixed effect, as well as a random slope (Barr et al., 2013). These models were then compared individually against the null model using the $\chi 2$ statistic. The results of these comparisons can be seen in Table 5.4 below.

Table 5.4.
Summary of Likelihood Ratio Tests on Predictor Variables on the Logit Mixed-effects Model of Pilot Study 1

Predictor Variable	$\chi 2$	df	p	C
Group	90.35	3	< .001***	.67
Case	1.65	7	.98	.66
Syntax	0.30	3	.96	.65
Grammaticality	7.23	3	$.06^{+}$	.67
Transfer	1.12	3	.77	.65

$^{+}$p<.1, *p<.05, **p<.01, ***p<.001

As can be seen in Table 5.4 above, only the fixed effect for group showed a significant relationship with the dependent variable. However, as the fixed effect of grammaticality appeared to be trending towards significance, this fixed effect was included in the combined model on an exploratory basis. Thus, a new model was constructed, which included group and grammaticality as fixed effects, crossed random intercepts for subjects and items, a random slope for grammaticality by participant, and a random slope for group by item. A likelihood ratio test indicated that this model was a significantly better fit for the data than was the original null model, χ^2 (7) = 101.55, $p < 0.001$, C =.71. Multicollinearity was formally assessed using Cohen's kappa. This resulted in a kappa value of 2.91 which, according to Baayen's (2008) guidelines, indicates a very low level of collinearity between the fixed effects in the model. A summary of this model can be found in Table 5.5 below.

Table 5.5.
Results for the Best-fitting Logit Mixed-effects Model Examining Performance in the GJT for Pilot Study 1

					Random Effects	
	Fixed Effects				By subject	By items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	-.02	.13	-.12	.90	.26	.51
Group	.12	.24	1.24	.62	–	.94
Gram	.23	.18	1.24	.22	.48	–
Group: Gram	-.03	.35	-.09	.93		

Note. Factors were coded using contrast coding[20], as follows: Group (−.5 = Control, .5 = Experimental), Grammaticality (-.5 ungrammatical, .5=grammatical), formula: "correct~group*grammaticality + (grammaticality|subject) + (group|item), family=binomial).

As can be seen in Table 5.5 above, neither the fixed effects of group, grammaticality, nor the interaction between group and grammaticality reached significance within this model. Therefore, it can be inferred that the performance of the experimental group did not differ significantly from that of the control group. In other words, the results of the logit mixed-effects analysis do not provide evidence of learning of the case marking system.

However, it should be mentioned that further breakdowns of the performance of the experimental group pointed towards an interesting finding. In particular, this analysis indicated that participants might have displayed a bias in their responses in the GJT that was not made apparent by the initial sub-analyses. One such pattern that emerged was that the participants displayed a clear bias in their responses towards

[20] As noted by Linck and Cunnings (2015), contrast coding aids in ensuring that the mixed-effects model compiles successfully and also helps to reduce multicollinearity between the predictor variables.

foreign words that they had previously encountered in the training phase. For *Transfer grammatical items*, namely test sentences with Czech words that were not included in the training set, participants classified only 34.58% (*SD* = 11.71) of the sentences correctly. By contrast, for *Trained grammatical items,* or sentences with Czech words that the participants had encountered in the training phase, participants classified 76.28% (*SD* = 15.75%) of the sentences accurately. These results suggest that the participants were more likely to judge a sentence as grammatical if the Czech word was familiar, and ungrammatical if the Czech word was unfamiliar. The same pattern held for ungrammatical items, where participants classified 66.99% (*SD* = 16.24) of transfer items correctly and 34.60% (*SD* = 15.40) of trained ungrammatical items correctly, again suggesting that participants were basing their judgements on the familiarity of the foreign word, rather than on the grammaticality of the case marker. In short, the results for the target case markers seem to have been skewed due to a shortcoming in the experimental design.

5.1.4. Interim Discussion

As noted above, there were two research questions for this study. Research question 1 examined the degree to which inflectional morphology can be acquired as a result of incidental exposure. Research Question 2 asked whether the knowledge gained as a result of this exposure is implicit or explicit in nature. Given the fact that no overall learning effect was demonstrated, it is assumed that no knowledge, neither implicit nor explicit, was acquired as a result of the training phase in Pilot Study 1. Therefore, the results of this experiment will not be discussed in the light of Research Question 2.

Regarding Research Question 1, the degree to which L2 inflectional morphology can be learned under incidental learning conditions, an analysis of the performance of the experimental participants did not reveal an overall learning effect. As noted above, further breakdowns of the performance of the experimental group revealed that their performance was highly skewed according to whether the test item contained a noun that was familiar; in other words, that had also been included in the training set. Taking the overall lack of a significant learning effect into account, it would appear that the training task might not have been sufficient to promote the learning of the target morphological structures. Of interest here is the experimental group's performance in the picture-matching task during the training phase of the experiment. As noted in the results section above, the experimental group demonstrated almost perfect accuracy when asked to match the picture to the meaning of the foreign word after listening to each sentence in the training set. This near ceiling-level performance suggests that this task met its original intention–each sentence was written so that an approximate meaning of the foreign word could be deduced from the context.

Despite this, when we consider that the experimental group's performance during the testing phase of the experiment was not significantly different from that predicted by chance, it would appear that the training phase of the experiment did not result in the participants developing knowledge of the underlying morphological rule system. However, the question of why this was the case remains. As noted in Section 4.1 above, inflectional morphology is an aspect of grammar that is notoriously difficult for L2 learners to acquire (DeKeyser, 2005). To account for this difficulty, a number of variables have been hypothesised to impact on the degree to which learners process, and successfully acquire, inflection morphology in an L2. For example, as noted in Section 4.3.3, Goldschneider and DeKeyser (2001) found that perceptual salience, semantic complexity, morphophonological regularity, syntactic category and frequency all play a role in acquisition. In the present experiment, three different case endings were used; nominative -a (/ə/), accusative -u (/u/) and instrumental –ou (/oʊ/).

Still more factors that have been argued to impact on the degree to which L2 learners notice and process morphological form were discussed by Gor (2010). This review highlighted that individual difference variables, properties of the L1, the properties of the L2, and the properties of the input itself, such as modality, all affect the degree to which the learners process inflected forms in the input. In the case of the present experiment, it seems plausible that the properties of the input might have played an important factor in the degree to which the participants needed to attend to the relevant features of the input. As noted in Section 4.3.1, theories of input processing (Van Patten, 2004b) hold that learners are unlikely to process morphological markers that have a redundant function and/or semantic mapping (Sagarra, 2008; Van Patten, 2004b). In other words, if participants are able to decipher the meaning of an utterance through lexical means, they are less likely to attend to grammatical markers that encode the same meaning. For example, if we analyse a sentence from the training set of Pilot Study 1, *Last summer the grass ate the koza in the field*, the verb eat typically entails an animate agent. Given that *grass* is inanimate, a participant would be able to surmise that *koza* is most likely the subject of the sentence. Thus, this utterance contains two redundant cues that provide information about the subject of the sentence: the inflectional marker –*a* and the verb "ate". This redundancy could then be interpreted as suggesting that the participants in Pilot Study 1 might not have noticed the case markers in the training phase of the experiment because they were able to comprehend the sentence via lexical means. Similar interpretations can be found in previous studies that examined the learning of L2 inflectional morphology. For example, Robinson (2002, 2005) noted that the higher performance on locative case markers, as opposed to ergative and noun-incorporation markers, could be explained in relation to the opacity of the form-meaning mapping carried by these markers.

In short, when taking into account the fact that the participants were able to complete the picture-matching task in the training phase with a high degree of accuracy, it appears that the training conditions did not necessitate that the participants attended to the morphological markers in order to complete the given task. Given theoretical arguments that noticing is a facilitative, if not essential, condition for learning to occur (Schmidt, 2001), future experimental designs would be strengthened by modifying the training conditions to provide more opportune conditions for noticing to take place. In previous studies of incidental learning (Rebuschat, 2008; Rebuschat & Williams, 2012), in which no learning was detected in initial experiments, the researchers were able to trigger development by modifying the training task to include elicited repetitions (asking participants to repeat the stimulus sentence aloud) in addition to making a judgement based on the content of the sentence.

In addition to the semantic redundancy present in the stimulus sentences, the syntactic complexity of the stimulus array may also have impacted on the degree to which the participants attended to the relevant morphological markers during the training phase of Pilot Study 1. Supporting this claim are the results from the verbal reports of Pilot Study 1, in which several participants commented that the "strange" word order made it more difficult to follow the overall meaning of the sentences. It would appear logical that more complex sentences place greater demands on attentional resources in order to comprehend them. This interpretation also appears to concur with points discussed above in relation to input processing, in particular to claims that learners prioritise meaning over grammatical form during input processing, as well as to arguments for attention being characterised as a limited-capacity system (Baddeley, 2012; see also Neumann, 1996; Schmidt, 2001; Tomlin & Villa, 1994 for a general discussion of the nature of the construct of attention). Following this line of argument, if the more complex, less "English-like" sentences were more difficult for the participants to understand, they would be unlikely to direct their attention towards the morphological markers. It would then stand to reason, in the case of the present study, that limiting the syntactic complexity of the sentences could potentially reduce the processing load during the training task, thus freeing up cognitive resources to attend to the relevant morphological markers.

Taking the points above into account, it appears that one way to move forward in the present line of investigation would be to modify the conditions of the training phase of the experiment with the aim of increasing opportunities for learners to attend to the relevant forms in the input. It would also seem likely that reducing the overall complexity of the syntactic patterns might aid in promoting noticing during the training phase of the experiment, and thus trigger learning.

5.2. Pilot Study 2

As with Pilot Study 1, Pilot Study 2 set out to address RQ 1 and RQ 2 in order to investigate the extent to which Czech morphology, specifically the nominative marker *–a*, the accusative marker *–u*, and the instrumental marker *–ou*, can be acquired under incidental learning conditions. As in Pilot Study 1, Pilot Study 2 followed the AGL paradigm to address the following research questions:

1). To what degree can L2 case markings be learned under incidental learning conditions?

2). What type of knowledge is acquired as a result of this exposure: implicit or explicit?

Pilot Study 2 was identical to Pilot Study 1, with the exception of three alterations that are outlined below.

5.2.1. Methods

5.2.1.1. Participants

Twenty-eight native speakers of English participated in Pilot Study 2. These participants were randomly assigned to experimental (*n* = 14; 9 female, 5 male) and control (*n* = 14; 8 female, 6 male) groups. The demographics of the participants were similar to those of the participants in Pilot Study 1. The ages of the participants (English L1) ranged from 18 to 24 years (*M* = 20.22, *SD* = 1.71).

The data collection for this experiment stopped short of the intended target (a minimum of 20 participants per group) due to the lack of a significant learning effect and small effect sizes.

5.2.1.2. Research Ethics

As with Pilot Study 1, this experiment was approved as exempt by the research support office of Lancaster University. Informed consent to take part in the experiment was sought from each individual participant prior to the start of the training phase of this study. Furthermore, the same ethical procedures were followed as described in Section 5.1.2.2, above.

5.2.1.3. Stimulus Material

Two modifications were made between Pilot Study 1 to 2 with regard to the stimulus materials for the training and testing sets. Firstly, in order to reduce the overall level of complexity of the stimulus material, the number of syntactic patterns for both the training and the testing sets was reduced from four to two, as can be seen in Table 5.6, below.

Table 5.6.
Templates, Sample Sentences, and Frequencies for the Two Syntactic Patterns Included in the Training and Testing Sets for Pilot Study 2

Pattern	*Template*	*Frequency*
Pattern 1	$[[AP]_{TEMP} > [NP]_{OBJ} > [VP] > [NP]_{SUBJ} > [PP]]$	
Nominative	*Last summer the grass ate the koza in the field.*	(24)
Accusative	*Last month the kasu opened Patrick with the key.*	(24)
Instrumental	*Some time ago John scared the child with a zrudou.*	(24)
Pattern 2	$[[AP]_{TEMP} > [NP]_{SUBJ} > [VP] > [NP]_{OBJ} > [PP]]$	
Nominative	*Last year the prodejna shipped goods to the shoppers.*	(24)
Accusative	*All week the builder took his vrtacku to work.*	(24)
Instrumental	*Today the wooden board cut he with a pilkou.*	(24)

It was hypothesised that limiting the syntactic structures in Pilot Study 2 to two patterns would help to reduce the cognitive demands of the stimulus domain, thus freeing up cognitive resources, which could then potentially be allocated to attend to the target morphological structures. In addition, a second modification was carried out with regard to the stimulus sentences in the testing set of Pilot Study 2. As noted in the discussion in Section 5.1.4, above, participants in Pilot Study 1 displayed some bias in their responses depending on whether or not the test item contained a familiar or an unfamiliar Czech noun. Because of this, it was apparent that participants were not basing their judgements on the grammaticality of a sentence, but rather on whether or not the sentence contained a foreign word that they had encountered before. To address this issue, the testing set of Pilot Study 2 was modified so that it included only novel Czech nouns, namely Czech nouns that were not used in the training set.

5.2.1.4. Procedure

The final alteration to Pilot Study 2 was that the procedure for the training phase was modified to include elicited imitations; in other words, participants were asked to repeat the entire sentence aloud prior to judging which of the two pictures more closely matched the meaning of the foreign word. This procedure can be seen in Figure 5E below.

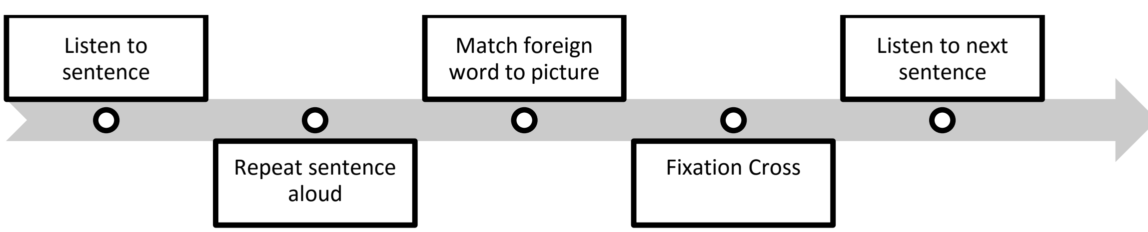


Figure 5E. *Procedure of the Training Phase in Pilot Study 2*

The rationale for including elicited imitations was that requiring participants to repeat the entire sentence aloud would necessitate attending to the entire sentence.

Thus, even if participants were to rely on lexis to interpret the overall meaning of the sentence, repeating the sentence aloud would help to ensure that the participants were attending to the surface features of the input, including the target morphological markers. In addition, as noted in the discussion of Pilot Study 1, when confronted with non-significant findings, previous incidental learning research (Rebuschat, 2008) managed to instigate learning by including elicited imitations in subsequent experimental designs. Examples of the slides used in the training phase of Pilot Study 2 can be found in Figure 5F below.

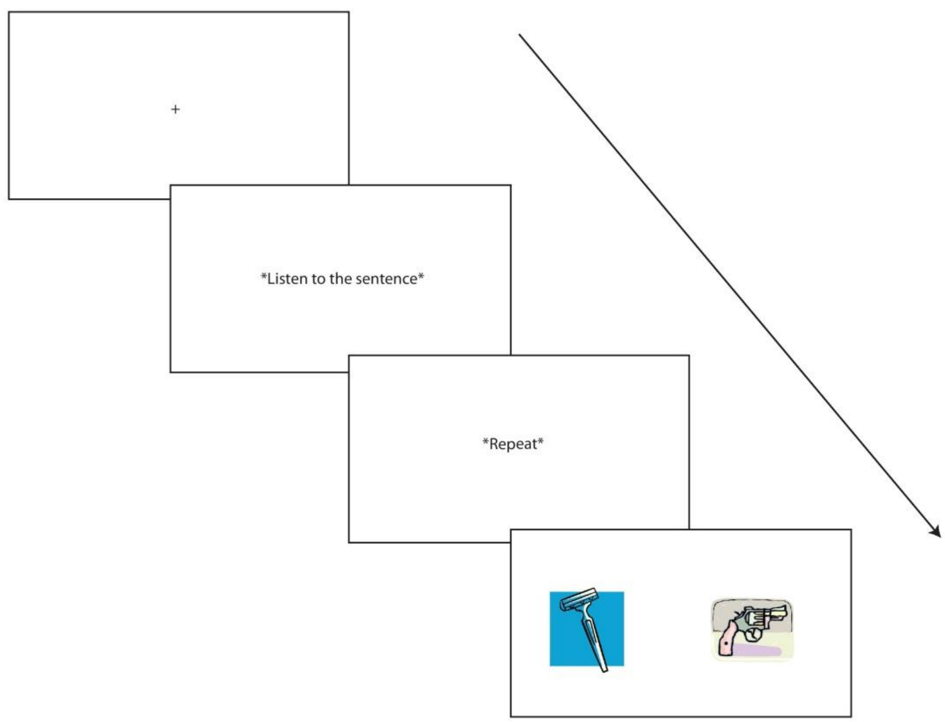


Figure 5F. *Example of the Slides Used in the Training Phase of Pilot Study 2*

5.2.1.5. Statistical Analyses.

Similar to Pilot Study 1, performance on a grammaticality judgement task was analysed in the first instance using mean accuracy rates and logit mixed-effects statistical modelling.

5.2.2. Results

5.2.2.1. Retrospective Verbal Reports

An analysis of the debriefing questionnaire revealed several interesting findings. Firstly, all the participants in the experimental group (14/14) reported noticing that the word order of the sentences was changing. However, in contrast to Pilot Study 1, these participants did not report that the changing syntax made it more difficult to comprehend the sentences, or to complete the training task. With particular reference to the case-marking system, none of the participants (0/14) reported becoming aware of the underlying rule system during the experiment. However, a majority (11/14) reported noticing that the endings of the foreign words were changing during the training phase of the study. When pushed to explain the rule underlying the changing case-endings, many of these participants (7/11) said that they assumed it was related to the gender of the noun. The remainder (4/11) reported that they thought that the foreign words might be changing, but were unsure as to whether they had misheard previous instances of the foreign words. When participants were asked what prompted them to notice the case markers, several of them stated that they only noticed the changing case-endings when the foreign words were repeated across different sentences in close proximity to each other, and several mentioned that they noticed the case markers on shorter foreign words, namely words consisting of two syllables (such as *myska*), as opposed to words with three syllables (such as *zahrada*), during the experiment.

5.2.2.2. Picture-matching task

An analysis of the participants' performance on the picture-matching task in the training phase of the experiment indicated that they were able to match the foreign word with its corresponding picture correctly and with great accuracy (*M* = 98.30%, *SD* = 1.14%). This result is similar to the results from Pilot Study 1, and it suggests that the participants were able to work out part of the intended meaning of the sentences in the training phase of the experiment. Furthermore, it was noted in the experimenter's report that the participants were able to repeat the sentences aloud successfully, despite the scrambled syntax and the inclusion of a foreign word. Taken together, these results suggest that the addition of the elicited imitations did not affect the degree to which the participants were able to complete the picture-matching task in the training phase of the experiment.

5.2.2.3. GJT

As was the case with Pilot Study 1, descriptive statistics were generated to compare the performance of the experimental and control groups. As in Pilot Study 1, the overall accuracy on the GJT of the experimental group (*M* = 51.34%, *SD* = 5.52%) was similar to that of the control group (*M* = 47.92%, *SD* = 7.30%). In addition, as can be seen in Table 5.7 below, the performance of the experimental and the control groups appeared similar across the different case markers and syntactic patterns, as well as for grammatical and ungrammatical test items.

Table 5.7.
Results of Experiment 2: Descriptive Statistics (%) and Results of Independent Sample T-tests for Experimental and Control Groups across Sub-components of the GJT

Item Type	*Descriptive Statistics (%)*		
Group	*M*	*SD*	*SE*
Nominative			
Experimental	47.22	15.83	4.23
Control	48.21	9.31	2.49
Accusative			
Experimental	53.57	16.02	4.28
Control	50.45	10.82	2.89
Instrumental			
Experimental	53.13	10.62	2.84
Control	45.09	9.55	2.55
Grammatical			
Experimental	51.79	9.90	2.65
Control	47.02	15.80	4.22
Ungrammatical			
Experimental	50.89	8.52	2.28
Control	48.81	11.37	3.04
Pattern 1			
Experimental	53.87	7.92	2.12
Control	49.40	9.92	2.65
Pattern 2			
Experimental	48.81	7.29	1.97
Control	46.43	8.15	2.18

As was the case in Pilot Study 1, a logit mixed-effects regression model (see Section 5.1.2 above for a discussion of mixed-effects modelling) was constructed using R (R Core Team, 2014) and lme4 (Bates et al., 2014) to examine whether any statistically significant differences existed between the performance of the experimental and of the control groups. Within this model, performance was modelled as a binary-outcome (correct versus incorrect), subjects and items were included as crossed random effects, and group (experimental versus control), case (nominative, accusative, instrumental), syntax (Type 1 versus Type 2) and grammaticality (grammatical versus ungrammatical test items) were analysed as fixed effects.

Initial analyses were carried out to determine which fixed effects had a significant relationship with the dependent variable, namely the performance in the GJT. To this end, a series of logit mixed-effects analyses were carried out. For these analyses, a null model was constructed, which analysed performance on the GJT as a binary outcome (correct versus incorrect) in terms of subject and item as crossed random intercepts. Additional models were then built for each fixed effect. Each of these models included the variable as a fixed effect, as well as a random slope (Barr et al., 2013). As was done with the previous analyses in this thesis, these models were then compared individually against the null model using the $\chi 2$ statistic. The results of these comparisons can be seen in Table 5.8 below.

Table 5.8.
Summary of Likelihood Ratio Tests on Predictor Variables in the GJT in Pilot Study 2

Predictor Variable	$\chi 2$	df	p	C
Group	215.14	3	< .001***	.81
Case	4.28	7	.75	.72
Syntax	0.19	3	.98	.70
Grammaticality	5.88	3	.12	.72

As can be seen in Table 5.8, only the fixed effect of group showed a significant relationship with the dependent variable. Therefore, a model that included group as a fixed effect and as a random slope by item was the best fit model for this dataset. Multicollinearity was formally assessed using Cohen's kappa. This resulted in a kappa value of 1.19 which, according to Baayen's (2008) guidelines, indicates a very low level of collinearity between the fixed effects in the model. However, as Table 5.9 below shows, a summary of this model indicates that the fixed effect of group did not reach significance. Thus, the results provide no evidence of a difference in performance between the experimental and the control groups, and therefore no evidence of the learning of the target morphological structures.

Table 5.9.
Results for the Best-fitting Logit Mixed-effects Model Examining Performance on the GJT for Pilot Study 2

					Random Effects	
	Fixed Effects				By subject	By items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	-.04	.12	-.31	.75	< .001	.74
Group	-.05	.08	-.62	.53	–	.51

Note. Factors were coded using contrast coding, as follows: Group (−.5 = Control, .5 = Experimental), Grammaticality (-.5 ungrammatical, .5=grammatical), formula: "correct~group*grammaticality+ (grammaticality|subject) + (group|item), family=binomial).

5.2.3. Discussion

Despite the various changes made to the stimulus materials and experimental procedure, the training phase did not result in a clear learning effect in Pilot Study 2. This nonsignificant result held true with regard to both the overall performance of the experimental group, and to their performance in the various sub-categories of the GJT.

Although a learning effect was not found in the present experiment, it is worth evaluating the effectiveness of the changes that were made between Pilot Studies 1 and 2. In particular, it is important to determine whether these changes were successful with regard to the original rationale(s) for making these modifications. As has been described above, there were two major changes to the training phase of the experiment from Pilot Study 1 to Pilot Study 2. First, in order to reduce the overall complexity of the stimulus domain, the number of syntactic patterns was reduced from four to two. The second change was that the procedure during the training phase was modified to include elicited imitations. This change was carried out under the assumption that asking the participants to repeat the sentences aloud would lead to more noticing of the target morphological markers.

Concerning the first change (reducing the number of syntactic structures), it is worth noting that all of the participants in Pilot Study 2 freely reported noticing the changing word order of the stimulus sentences. However, unlike Pilot Study 1 in which participants reported having difficulty following the meaning of the sentences, the participants in Pilot Study 2 did not report such difficulties. Based on these reports, reducing the complexity of the stimulus array could be seen as a successful alteration, due to the fact that reducing the syntactic complexity appears to have reduced the difficulty that the participants had in comprehending the stimuli.

Although syntactic complexity undoubtedly contributed to the relative difficulty of the training task in Pilot Study 2, another variable that may have impacted on the findings concerns the number of times each syntactic pattern was repeated throughout the training phase. It should be noted that Pilot Study 1 consisted of four syntactic patterns, which were repeated 36 times each for a total of 144 exposures. In order to keep the total number of exposures constant across Experiments 1 and 2, the two syntactic patterns in Pilot Study 2 were each repeated within 72 sentences. Given the argument that repetition plays an important role in SLA (N. Ellis, 2006), it seems plausible that the repetition here might also have had some impact on the degree of difficulty of the training task. In particular, repetition might have led to the participants growing more accustomed to these patterns over the course of the training phase.

Concerning the second modification (adding elicited imitations to the training set), it is less certain whether this change was successful as per the original assumptions. As noted above, the experimenter's report noted that the participants were able to repeat the sentences aloud, including the inflected foreign word. When we compare the results of the retrospective verbal reports, it appears that a higher proportion of participants in Pilot Study 2 (11/14) reported noticing the changing inflections than was the case with the proportion of participants in Pilot Study 1(12/28). It is important to stress, however, that this finding should not be interpreted as direct evidence that the elicited imitations led to more noticing on the part of the participants, particularly given the limitations of retrospective verbal reports. As noted in Section 3.5.2 above, retrospective verbal reports have frequently been criticised as an insensitive measure of awareness, especially with regard to concurrent awareness during the training phase of an experiment. Given the multiple changes between Experiments 1 and 2, it is not possible to say which change, or combination of changes, led to more noticing being reported in the retrospective verbal reports. However, it should be stressed that the data here suggest that, although the changes from Pilot Study 1 to Pilot Study 2 did not lead to a significant learning effect in the GJT, these changes do appear to have fostered more reported noticing (Schmidt, 1990) of the target morphological markers during the training phase of the experiment.

Despite the evidence above, which suggests that more noticing of the target structures took place in Pilot Study 2, it appears that the training conditions were not sufficient to spark a learning effect amongst the participants. At this point, it is worth returning to the retrospective verbal reports for guidance as to how future experiments might be modified to promote learning. As noted in the results section above, several of the participants commented during the debriefing session that they noticed the endings of the shorter foreign words, such as *kocka* and *myska*, during the training phase. Although the length of the foreign word was not a variable that was controlled for in either Pilot Study 1 or 2, previous research into the acquisition of inflectional morphology (Brooks et al., 2006) has indicated that word length does impact on the acquisition of word form. This impact is partly attributable to the fact that longer words place more demands on phonological short-term memory (N. Ellis, 1996; Gupta, 2003). Thus, controlling for word length might be an avenue to explore in future experiments, in that using only shorter words might free up additional cognitive resources that could, in theory, be allocated towards more in-depth processing of the target morphological markers.

In addition to word length, some participants also reported that they only noticed the endings when the base forms of the foreign words were repeated with a different case marker. It seems then, as per one of this study's original assumptions, that the repetition of a foreign word across different morphological categories might promote noticing of the changing case ending. Furthermore, future experiments might benefit in additional ways from manipulating the training phase in light of the positive effects of repetition. For example, the order of the sentences in the training set could be pseudo-randomised, so that the three instances of each foreign word occur within close proximity. Another possibility might be to insert more repetition directly into the training set, for instance by reducing the total number of sentences, but repeating each of these sentences several times in the exposure phase.

5.3. Summary of Pilot Studies 1 and 2

In summary, the results of Pilot Study 1 and Pilot Study 2 serve to illustrate that the learning of a semi-artificial system under incidental learning conditions might not be a straightforward matter; no learning effect was observed in either study, despite the fact that the morphological system used in both studies represented a greatly simplified version of a natural linguistic system. The linguistic system in Pilot Study 1 consisted of three morphological patterns that were used individually within sentences that followed one of four syntactic patterns. The training task in this experiment was designed so that participants' attention was focused on the meaning of the foreign word, rather than on the grammatical form. Participants listened to a sentence, and then indicated which of two pictures best matched the meaning of the foreign word in the given sentence. The results of the GJT indicate that the training conditions were not sufficient to promote learning of the morphological system. Pilot Study 2 simplified the linguistic system of Pilot Study 1 by reducing the number of syntactic patterns from four to two. In addition, the training phase of Pilot Study 2 was modified so that the participants listened to a sentence, repeated the sentence aloud, and then indicated which of two pictures best matched the meaning of the foreign word in the sentence. It was hypothesised that these two changes would result in more conscious noticing, and subsequent learning, of the target morphological markers. This was, however, not the case.

It is clear at this stage that changes need to be made to future experiments if these experiments are to be successful in promoting the learning of the semi-artificial system utilised here. Based on the evidence collected via the retrospective verbal reports, it appears that one way forward is to further simplify the linguistic system in this study, as well as to make additional changes to the training task to promote the degree to which participants attend to the case endings present in the input.

CHAPTER 6

EXPERIMENTS 1 AND 2

This chapter presents two of the experiments of this thesis, which are referred to henceforth as Experiment 1 and Experiment 2. As with Pilot Studies 1 and 2, both Experiment 1 and 2 followed the AGL paradigm to address the following research questions:

1). To what degree can L2 case markings be learned under incidental learning conditions?

2). What type of knowledge is acquired as a result of this exposure; implicit or explicit?

This chapter is organised as follows. First, the chapter begins by providing a brief rationale for Experiment 1, which is followed by a more detailed description of the population, methods, materials, and experimental procedure. The results of Experiment 1 are then presented, followed by an interim discussion and summary. This general outline is then repeated for Experiment 2. This chapter concludes with a brief summary of both experiments, including a critical evaluation of the research design utilised within them, as well as recommendations for future experiments.

6.1. Experiment 1

As in Pilot Studies 1 and 2, Experiment 1 set out to address RQ 1 and RQ 2 in order to investigate the degree to which Czech morphology, specifically the nominative marker *–a* and the accusative marker *–u*, can be acquired under incidental learning conditions. Experiment 1 was identical to Pilot Study 2, but for three alterations. The first alteration was that the number of morphological categories was reduced from three to two. The second was that the number of exemplars, namely unique sentences, was decreased from 144 to 48. Each of these exemplars was then repeated three times to keep the total amount of exposure constant across the experiments. Finally, concerning the task in the training phase, participants were asked to repeat the sentence aloud, and to then repeat the foreign word in isolation before finishing the sentence-completion task. These changes between the experiments are explained more thoroughly in the sections that follow, and they are summarised in Table 6.1 below.

Table 6.1.

Summary of Changes to Training and Testing Phases between Pilot Study 1, Pilot Study 2, and Experiment 1

	Training Phase						Testing Phase
	Exemplars	Repetition	Total Exposure	Case markers	Syntactic Patterns	Training Task(s)	Testing Instruments
Pilot Study 1	144	0	144 sentences (auditory only)	Nominative Accusative Instrumental	4	1. Picture-matching	1. 48-item auditory GJT (24 trained, 24 transfer) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire
Pilot Study 2	144	0	144 sentences (auditory only)	Nominative Accusative Instrumental	2	1. Picture-matching 2. Elicited imitations (repeat entire sentence aloud)	1. 48-item auditory GJT (48 transfer items) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire
Experiment 1	48	3x	144 sentences (auditory only)	Nominative Accusative	2	1. Picture-matching, 2. Elicited imitations (repeat sentence), 3. Elicited imitations (repeat foreign word)	1. 48-item auditory GJT (48 transfer items) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire

6.1.1. Method

6.1.1.1. Participants

Fifty-one undergraduate native speakers of English at a university in the United Kingdom volunteered to take part in this study. Nine of these volunteers were excluded from the present study on the basis of having previous experience with a language with a rich morphological system, such as Latin. The remaining 42 participants (26 female, 16 male) were randomly assigned to an experimental (n = 21; 14 female, 7 male) or a control (n = 21; 12 female, 9 male) group. None of the participants selected for this study majored in linguistics or foreign languages. The ages of the participants ranged from 18 to 25 years (M = 20.03, SD = 1.86).

6.1.1.2. Research Ethics

As in Pilot Studies 1 and 2, this experiment was approved as exempt by the research support office of Lancaster University. For each participant, informed consent was sought prior to beginning the experiment. Each participant was allowed to read through the information sheet describing the study, and he or she was given an opportunity to ask questions or voice any worries or concerns about the experiment. Furthermore, as in the previous pilot studies, it was stressed that participation in this study was voluntary, and that participants could withdraw at any point.

It should again be noted that the information sheet for this study disguised the true purpose of this experiment. As with Pilot Studies 1 and 2, Experiment 1 was presented as a study on learning foreign language vocabulary. Again, this deception was necessary to ensure the construct validity of the incidental learning conditions. It was only at the end of the debriefing session that participants were fully informed of the focus of this experiment, and were given an additional opportunity to ask questions, voice concerns, or withdraw their consent.

6.1.1.3. Stimulus Material

The first two alterations to Experiment 1 were made to further reduce the overall complexity of the training and testing sets. First, the number of morphological categories was reduced from three (nominative, accusative, instrumental) to only two (nominative, accusative) in both the training and testing sets (see Table 6.2, below, for sample sentences).

Table 6.2.
Template, Sample Sentences, and Frequencies for the Syntactic Patterns and Morphological Categories in Experiment 1

Pattern	*Template*	*Frequency*
Pattern 1	$[[AP]_{TEMP} > [NP]_{OBJ} > [VP] > [NP]_{SUBJ} > [PP]]$	
Nominative	*Last summer the grass ate the koza in the field.*	(12)
Accusative	*Last month the kasu opened Patrick with the key.*	(12)
Pattern 2	$[[AP]_{TEMP} > [NP]_{SUBJ} > [VP] > [NP]_{OBJ} > [PP]]$	
Nominative	*Last year the prodejna shipped goods to the shoppers.*	(12)
Accusative	*All week the builder took his vrtacku to work.*	(12)

In addition, the training set was halved from 48 nouns to include only 24 nouns. Only nouns with two syllables were retained. These 24 nouns were repeated three times for both nominative and accusative cases across three training blocks for a total of 144 total stimulus sentences (72 nominative and 72 instrumental). This maintained the same total amount of input as in Experiments 1 and 2, but represented an increase in exposure for each morphological category. The complete stimulus set can be found in Appendix C.

6.1.1.4. Procedure

As with Pilot Studies 1 and 2, Experiment 1 took place in a private office on campus. The training and testing phases of the experiment were delivered via a Dell Thinkpad T410 laptop computer, using the stimulus presentation software Superlab 4.5 (Cedrus Corp, San Pedro, CA), together with a Cedrus model RB-834 response pad. The benefit of the response pad for the current experiment is with regard to the ease of experimentation. In comparison to using a keyboard, the response pad has large keys with good spatial orientation, thus making it easier for participants to identify and press the appropriate key during the experiment (see Figure 6A, below, for a depiction of the response pad as used in the present thesis). All audio for the experiment was played through a set of Audio-Technica headphones with noise-cancellation functionality. Following the testing phase, participants completed a short debriefing questionnaire, followed by an oral interview. All participants in this study, that is in both the experimental and control groups, gave their informed consent prior to the start of the experiment.

Figure 6A. *Cedrus model RB-834 Response Pad*

Training Phase

The final alteration concerned the training procedure, which was modified from Experiment 2 in that the participants had to repeat the foreign word in isolation, in addition to repeating the entire sentence. A visual representation of this procedure can be seen in Figure 6B below.

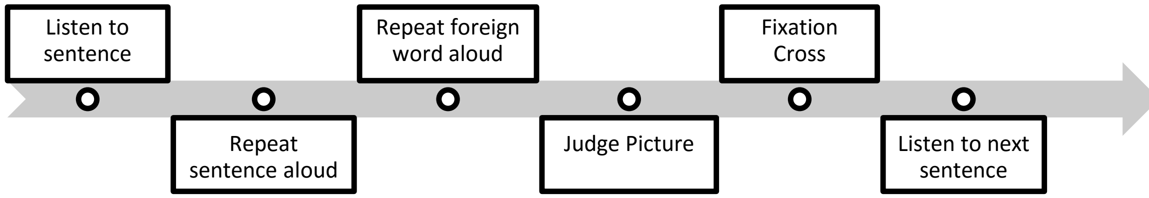


Figure 6B. *Training procedure in Experiment 1.*

As Figure 6B above illustrates, the training procedure required the participants to listen to the sentence (for example, *The britva cut David's face at the sink last night)*, to repeat the entire sentence aloud (*The britva cut David's face at the sink last night)*, and to then repeat the foreign word in isolation (*britva*), before judging which of the two pictures on the monitor (a razor or a gun) best matched the meaning of the foreign word. The entire training phase took, on average, about 30 minutes to complete. The slides used for Experiment 1 were identical to those used in Pilot Study 2. Examples of these slides can be seen in Figure 6C, below.

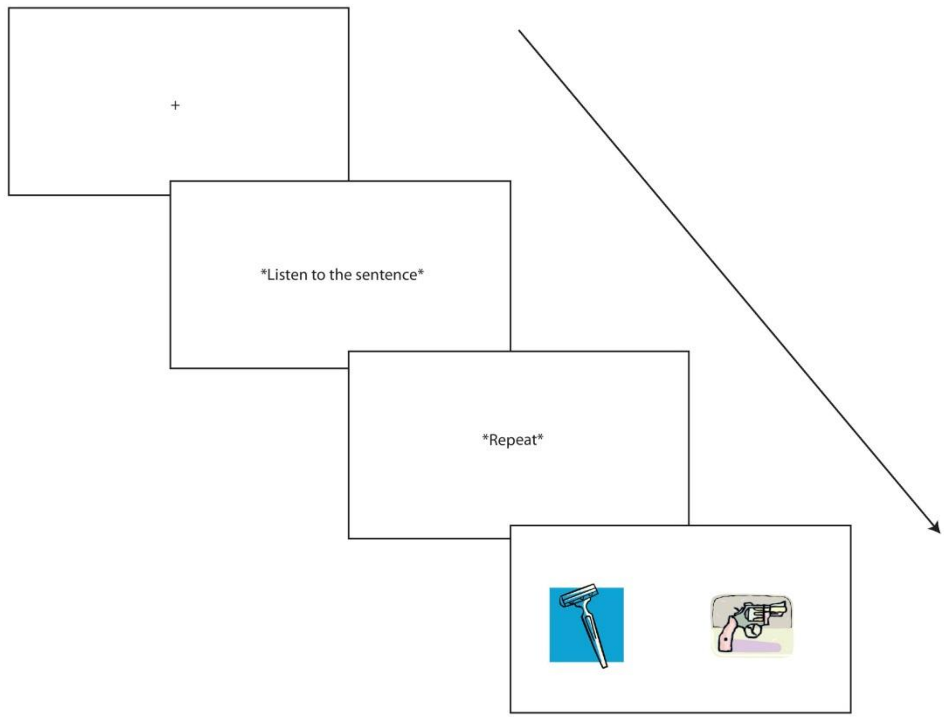


Figure 6C. *Example of the Slides Used in the Training Phase of Experiment 1*

6.1.1.5. Statistical Analyses

As in Pilot Studies 1 and 2, performance on the grammaticality judgement task in Experiment 1 was analysed using a logit mixed-effects model (Jaeger, 2008; Linck & Cunnings, 2015). The procedure for building this model was identical to the procedure in Pilot Studies 1 and 2. In building the model, statistical significance for the overall model fit was determined using the χ^2 statistic of likelihood ratio tests. Following the example provided by Baayen (2008), effect sizes for the logit-mixed effects model were calculated using the *C* index of concordance and the Hmisc package (Harrell & Dupont, 2015). C index values range from .5 to 1, with .5 representing no fit and 1 representing a perfect fit of the data. As noted in the preceding chapter, a C-index of .7 is interpreted as a moderate fit, .8 and above as a good fit, and .9 and above as an excellent fit for the data (Gries, 2013). Statistical significance for any interaction effects were calculated using post-hoc Tukey tests. Effect sizes for these tests were calculated using Cohen's *d*. Following Plonsky and Oswald's (2014) recommendations, Cohen's *d* of .40 is considered small, .70 medium, and 1.00 and higher as a large effect size.

6.1.2. Results of Experiment 1

6.1.2.1. Retrospective Verbal Reports

An analysis of retrospective verbal reports indicated that all participants (21/21) reported noticing the morphological inflections at the end of the foreign words during the training phase of the experiment. Comments such as "the foreign words always ended in vowels" or, more specifically, "all the words ended in *-u* or *-a*" as well as "the endings of the words changed" were taken as evidence that the learners had attended to and subsequently become aware of the surface features of the morphological markers. Furthermore, these comments also suggested that the learners had developed some awareness beyond the surface forms – specifically, that these inflections were part of a pattern or behaved in a certain way. Despite this, none of the participants were able to demonstrate an understanding of the underlying rule system by verbalising the rules governing the use of inflections. When prompted to guess, 13 participants stated that the affixes might have represented a noun class, such as gender. Two other subjects mentioned that they thought that the inflection might be connected in some way to the position of the foreign word in the sentence, but could not explain the nature of this connection. At the very end of the experiment, when the rules were explained to the participants, none of the participants claimed that they had thought of the rules at any point during the experiment. In summary, the analysis of the retrospective verbal reports thus suggests that participants exhibited awareness at the level of noticing, but below the threshold of verbalisation (Schmidt, 1990). In addition, as in Pilot Studies 1 and 2, all of the participants (21/21) commented on the changing word order of the stimulus sentences. As in Pilot Study 1, several (5/21) of the participants noted that the word order was "strange" and that the unnatural word order made it more difficult to follow the meaning of these sentences at the start of the experiment.

6.1.2.2. Picture-matching task

As in Experiments 1 and 2, the participants were able to complete the picture-matching task in the training phase with a near perfect level of accuracy (*M* = 99.08%, *SD* =0.82%). This suggests that the participants' attention was oriented towards the training task in that they were able to follow general meaning of the sentences in the training phase of the experiment.

6.1.2.3. GJT

As was the case in Pilot Studies 1 and 2, descriptive statistics were generated to analyse the performance of the experimental and control groups in the GJT. In contrast to the previous experiments, the results revealed that the experimental group (*M* = 55.44%, *SD*= 7.00%) appeared to outperform the control group slightly (*M* = 49.71%, *SD* = 5.80%). Furthermore, as can be seen in Table 6.3 below, the experimental group appeared to outperform the control group across a number of sub-categories of the GJT, including for items targeting the accusative case, and for items that followed an O-V-S word order.

Table 6.3.
Descriptive Statistics (%) for Experimental (n=21) and Control (n=21) Groups across Grammatical, Ungrammatical, Nominative, Accusative, O-V-S and S-V-O Test Items

Item Type	*Descriptive Statistics (%)*		
Group	*M*	*SD*	*SE*
Grammatical			
Experimental	57.14	9.94	2.17
Control	51.79	14.54	3.17
Ungrammatical			
Experimental	54.76	10.63	2.32
Control	47.52	9.58	2.09
Nominative			
Experimental	50.89	10.13	2.21
Control	48.02	8.53	1.86
Accusative			
Experimental	61.01	14.10	2.40
Control	52.38	8.50	1.85
Type 1 S-V-0			
Experimental	54.76	11.84	2.58
Control	49.40	9.66	2.11
Type 2 0-V-S			
Experimental	57.14	10.88	2.37
Control	50.00	8.62	1.88

In order to determine whether these apparent differences in performance between the control and experimental groups were statistically significant, a logit mixed-effects regression model (Jaeger, 2008) was carried out using R (R Core Team, 2014) and lme4 (Bates et al., 2014). As noted, this model was constructed following the general procedure used within Pilot Studies 1 and 2. First, analyses were carried out to determine which fixed effects had a significant relationship with the dependent variable, namely performance on the GJT. To this end, a series of logit mixed-effects models were built. For these analyses, a null model was constructed, which analysed performance in the GJT as a binary outcome in terms of subject and item as crossed random intercepts. Additional models were then built for each fixed effect. Each of these models included the variable, such as case, as a fixed effect, as well as a random slope (Barr et al., 2013). As was done with the previous analyses in this thesis, these models were then compared individually against the null model using the $\chi 2$ statistic. If the model that included the fixed effect was significant against the null model, this result was then interpreted to indicate that the fixed effect in question had a significant relationship with the dependent variable, and should be included in subsequent analyses. If the result was non-significant, this was interpreted as there being no significant relationship, and this fixed effect could be excluded from the models that followed. The results of these comparisons can be seen in Table 6.4 below.

Table 6.4.
Summary of Likelihood Ratio Tests on Predictor Variables in the GJT

Predictor Variable	χ^2	df	p	C
Group	205.05	3	<.001***	.79
Case	10.17	3	.02*	.67
Syntax	0.65	3	.88	.64
Grammaticality	4.01	3	.26	.67

As can be seen in Table 6.4 above, the results of these comparisons indicated that the fixed effects of case (nominative versus accusative) and group (experimental versus control) showed a significant relationship with performance in the GJT. Therefore, a new model that included both of these variables as fixed effects was constructed. In addition, following the maximal model as suggested by Barr et al. (2012), the fixed effect of case was included as a random slope by participant, and group as a random slope by item. A maximum likelihood ratio test indicated that this model was significant against the null model, χ^2 (7) = 218.22, p < .001, C=.80. Further comparisons indicated that the inclusion of both case and awareness as fixed effects was warranted in the model, and the removal of either fixed effect did not lead to an improved model fit. Multicollinearity within this model was formally assessed using Cohen's kappa. This resulted in a kappa value of 6.88 which, according to Baayen's (2008) guidelines, indicates a reasonable level of collinearity between the fixed effects in the model. Thus, this model of the data from the GJT is henceforth referred to as the best-fit model, the results of which are presented in Table 6.5 below.

Table 6.5.
Results for the Best-fitting Logit Mixed-effects Model Examining Performance in the GJT

	Fixed Effects				*Random Effects*	
					By subject	By items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	0.75	0.18	4.05	<.001***	.19	2.81
Group	-0.40	0.11	-3.61	<.001***	–	.13
Case	1.08	0.37	2.92	<.01**	.55	–
Group: Case	-0.51	0.22	-2.31	.02*		

Note. Factors were coded using contrast coding, as follows: Case (−.5 = Nominative, .5 = Accusative), Group (-.5 control, .5=experimental), formula: "correct~case*group + (case|subject) + (group|item), family=binomial".

As noted in the table above, the fixed effect of group was significant, which indicates that the performance of the experimental group was significant against the performance of the control group. In addition, the fixed effect of case was significant. This indicates that the overall performance of both the experimental and control groups differed significantly with regard to nominative versus accusative items. However, there was also a significant interaction between the fixed effects of group and case. This points towards differences in performance between the control and experimental groups on either nominative or accusative items. Looking at the descriptive statistics in Table 6.6 below, there is a clear difference in the performance of the experimental group in terms of accusative (~60% performance) versus nominative (~50% performance) items. By contrast, the control group's performance regarding accusative items (~52%) appears similar to its performance for nominative items (~48%). This suggests that the significant finding for "case" was driven by the performance of the experimental group for accusative items. To analyse whether these differences were statistically significant, post-hoc Tukey tests were calculated using the GLHT command of the R package "multcomp" (Hothorn et al., 2015). These results revealed that the performance of the experimental group was significant against the performance of the control group for accusative test items, but not for nominative items. These results can also be seen in Table 6.6 below.

Table 6.6.
Descriptive Statistics (%) and Results of Post-hoc Tukey Tests for Experimental (n=21) and Control (n=21) Groups across Nominative and Accusative Items

Item Type	*Descriptive Statistics (%)*			Results of Post-hoc Tukey Tests				
Group	*M*	*SD*	*SE*	*B*	*SE*	*z*	*p*	*d*
Nominative								
Experimental	50.89	10.13	2.21	-0.15	0.16	-0.94	0.78	0.31
Control	48.02	8.53	1.86					
Accusative								
Experimental	61.01	14.10	2.40	-.66	0.16	4.14	<.001***	0.66
Control	52.38	8.50	1.85					

6.1.2.4. Subjective Measures of Awareness

As Experiment 1 resulted in a significant learning effect, it is possible to analyse the results of the subjective measures of awareness. As noted in Chapter 3, subjective measures of awareness are designed to determine whether resulting knowledge is implicit or explicit in nature. These were not discussed in Pilot Studies 1 and 2 due to the lack of a learning effect within these two experiments. Recall that subjective measures of awareness consist of two separate instruments: source attributions and confidence ratings (see Section 3.5.3 for a discussion). Source attributions ask participants to indicate the basis for each of their decisions in a test (such as a GJT). In the case of the present experiment, the participants had to choose between *guess*, *intuition*, *memory*, or *rule* for each of their decisions in the GJT. Confidence ratings asked participants to indicate how confident they were regarding their decision. In Experiment 1, participants had to choose between *no confidence*, *somewhat confident*, *very confident*, or *absolutely certain*. Two separate criteria for interpreting the results of subjective measures of awareness are argued to be indicative of implicit knowledge. The first of these criteria, the *guessing criterion*, is when there is a performance that is significant against chance levels when attributing decisions to guess or intuition (Dienes et al., 1995). This analysis is typically carried out using one-sample *t* tests, comparing accuracy against chance level performance (50%). The second, *the zero-correlation criterion*, is when there is no relationship between accuracy and the level of confidence. In other words, if knowledge is explicit, we would expect that accuracy would increase when participants report higher levels of confidence regarding their decisions. This analysis is commonly carried out by including confidence as a predictor variable within the logit mixed-effects model (see Rausch, Müller, & Zehetleitner, in press; Sandberg, Bibby, and Overgaard, 2013; Wierzchoń, Asanowicz, Paulewicz, & Cleeremans, 2012, for examples).

Source attributions. As can be seen in Table 6.7 below, an analysis of the source attribution data revealed that participants relied largely on their intuition rather than on explicit rule knowledge in their judgements. In addition, one-sample *t* tests revealed that the accuracy of the participants' judgements was only significantly above chance levels when they indicated that they had made their judgements based on intuition. This means that the *guessing criterion* of implicit structural knowledge was not met. However, there are two pieces of evidence that lend support to the claim that the participants developed some unconscious grammatical knowledge. Firstly, the performance of the experimental group was significant against chance on intuition judgements, which has been argued to be reflective of conscious judgement knowledge combined with some unconscious structural knowledge (Dienes, 2010, 2013). In other words, significant performance when attributing decisions to intuition points towards the fact that participants did not know why a particular item on the GJT was correct or incorrect (unconscious structural knowledge), yet they had an intuition or a feeling regarding whether the items were correct or incorrect; in other words, conscious judgement knowledge (see Section 3.5.3 for a discussion). This is in contrast to participants' nonsignificant performance when attributing decisions to *rule* – thus providing evidence that participants had not acquired explicit structural knowledge of the underlying rule system.

Table 6.7.
Accuracy (%) and Number of Responses from Experimental Group Participants across Confidence Ratings and Source Attributions

	Accuracy (%)	Number	*p*-value
Confidence Rating			
No confidence	51.9	155	.70
Somewhat confident	56.3	501	.02
Very confident	57.1	310	.04
Absolutely certain	59.3	42	.35
	Accuracy (%)	Number	*p*-value
Source Attribution			
Guess	50.5	150	.92
Intuition	56.3	511	.02
Memory	57.6	184	.08
Rule	58.8	163	.06

Confidence ratings. As shown in Table 6.7 above, one-sample *t* tests indicated that the participants' performance was significantly better than chance levels when indicating that they were somewhat confident or very confident about their decisions. However, as noted in Section 3.5.3, the criterion for implicit knowledge as measured by confidence ratings is the zero-correlation criterion (Dienes & Scott, 2005), which holds that knowledge can be considered implicit if there is no relationship between reported levels of confidence and accuracy (for example, in a GJT). To determine if the zero-correlation criterion was met, a logit mixed-effects model with crossed random effects of subjects and items (Jaeger, 2008) was built in order to analyse if there was any underlying interaction between confidence and accuracy. This model was implemented using R (R Core Team, 2014) and lme4 (Bates et al., 2014) to examine the relationship between the level of confidence (no confidence versus somewhat confident versus very confident versus absolutely certain) and performance. Within this model, performance was modelled as a binary-outcome, 'confidence' as a fixed-effect, and 'Item' and 'Subject' as crossed random intercepts with random slopes. First, a null-model that incorporated only the random effects for the intercepts of "Item" and "Subject" was computed. Following this, a second model was calculated, which included "confidence" as a fixed effect. A likelihood ratio test was carried out to compare the null model to the model incorporating the effects of confidence. The results of this test did not reveal a relationship between confidence and accuracy, $\chi2$ (1) = 0.60, p =.44, C=.59. This result satisfies the requirements for the *zero-correlation criterion* (Dienes & Scott, 2005), and provides further evidence that the participants' judgement knowledge of the target case markers was, at least partially, unconscious in nature.

6.1.3. Interim Discussion of Experiment 1

The results of this experiment demonstrated that participants can develop some knowledge of L2 case markings after a minimal amount of incidental exposure. Concerning the relationship between learning and awareness in this study, an analysis of the retrospective verbal reports indicates that this learning effect was realised in the absence of verbalisable rule knowledge, but in the presence of low levels of awareness; in this case, the participants reported becoming aware of the surface features of the inflected forms during the training phase of the experiment. Furthermore, traditional analyses of subjective measures of awareness (guessing criterion, zero-correlation criterion) suggest that the exposure phase had resulted in some implicit knowledge of L2 morphology amongst the experimental subjects.

Research Question 1: To what extent can L2 case marking be acquired under incidental learning conditions?

One finding of particular interest was that the performance of the experimental group was significantly better than that of the control group for accusative case items, but not for nominative items. As noted above, this differential performance accords with the findings of Robinson’s (2002, 2005) study, which reported improved performance only for the locative marker, but not for the noun-incorporation and ergative markers in the miniature Samoan system he utilised to examine the acquisition of L2 morphology. Robinson (2002) speculated that the superior performance and concomitant awareness of the form-function mapping associated with the locative marker could be attributed to the amount of processing required by the locative training items. As Robinson (2002) noted, in order to answer the comprehension question related to the locative case items, participants had to attend to the information conveyed by the case marker. On the other hand, comprehension questions targeting the ergative case items, which showed the lowest levels of learning, could be solved without attending to the ergative marker due to the opacity and communicative redundancy of this construction. Unlike in Robinson's research, a differential amount of semantic processing does not seem to offer an explanation of the patterns observed in the present study. Here, it does not appear that the participants necessarily needed to attend to the semantic information conveyed by either the nominative or the accusative case markers during training. The training task required participants to match the meaning of the foreign word to one of two pictures, which could, in theory, be performed by relying solely on the lexical clues available within the sentences.

What, then, could account for the different performance for the nominative versus the accusative items? A possible explanation could be attributed to what has been labelled in cognitive psychology literature as the “bizarreness effect” (Riefer & Rouder, 1992); specifically, the finding that distinctive stimuli are better recalled than are more common stimuli. In the case of the present study, this effect could potentially be connected with the phonological characteristics of the morphemes used to mark the two cases. The phoneme associated with the nominative marker, *–a* (/a/), was probably perceived by English speakers as more natural in a word-final position than the phoneme used to realise the accusative marker, *–u* (/u/), given the phonotactic rules of the English language. Thus, the ‘oddity’ of the accusative morpheme /u/ might have resulted in increased salience, leading to more noticing of the accusative marker during the training phrase. This interpretation is also lent some support by the fact that participants in the experimental group were able to perform with more accuracy than were their counterparts in the control group for items with an O-V-S word order (a non-typical English word order), but not for items with an S-V-O word order (the prototypical word order in English). Although this difference did not reach significance within the mixed-effects model, it seems possible that the 'foreign-ness' of the O-V-S items might have resulted in increased attention during exposure.

Finally, given the multiple changes to the training procedures and stimulus materials between Pilot Study 2 and Experiment 1, it is not possible to pinpoint which component, or which combination thereof, is responsible for the incidental learning that was observed in this study. It seems likely that the repetition of the exemplars in the training set (both as part of the sentence and individually), and the elicited imitations as part of the training condition, were likely to have played a role in achieving a learning effect, given the role that frequency arguably plays in L2 acquisition (N. Ellis 2002, 2005, 2008). As previously discussed, the procedure of repeating the sentence aloud has also been used in previous experiments operating within the AGL paradigm (see, for example, Rebuschat & Williams, 2012). It could be argued that repetition led to an increased amount of rehearsal within short-term memory, creating additional opportunities for noticing to occur at the time of encoding.

One criticism that could be levied against the results of this experiment is that the performance of the experimental group was only slightly above chance levels (approximately 56%); thus, the advantage observed for the experimental group could possibly have been the result of a sampling error. While this criticism is best addressed through replication, these results should not be surprising given the difficulty in acquiring inflectional morphology (Larsen-Freeman, 2010), as well as the minimal amount of exposure the participants received in the present study (approximately 30 minutes). Furthermore, it is worth comparing the results of this experiment to those of previous research operating within the AGL paradigm, as well as within SLA. Artificial grammar learning studies typically result in a 55 to 80 per cent performance on grammaticality judgement tests (DeKeyser, 2003). Also, the results obtained here appear similar to those in Grey et al. (2014), the study that is most directly comparable to the current research. In Grey et al., the mean accuracy (*M* = 56.3%) in the delayed picture-matching task roughly corresponds to the mean average observed in the GJT in the present study (*M* = 55.4%). Although caution needs to be exercised when making comparisons across studies due to differences in their designs, participants and contexts, these trends indicate that the findings of the present study are not atypical within the AGL paradigm, nor in SLA research investigating the acquisition of L2 case marking following incidental exposure.

Research Question 2: What type of knowledge is acquired:implicit or explicit?

With regard to the second research question, the results of the retrospective verbal reports indicated that none of the participants had become aware of the morphological rule system. The lack of verbalisable knowledge shows that subjects can develop knowledge of L2 morphology without being able to account for their performance verbally. Although none of the participants in this study were able to verbalise the underlying rule system, all of them reported noticing, or becoming aware of, the changing inflections during the training phase of the experiment. This report of noticing is indicative of a low level of awareness of the target linguistic forms (Schmidt, 1990).

While the results of the retrospective verbal reports point to a low level of awareness on the part of the learners, the results of the subjective measures of awareness paint a slightly more complex picture regarding the nature of the knowledge acquired in this experiment. Firstly, concerning the confidence ratings, the immediate results suggest that participants developed both conscious and unconscious knowledge; in other words, the knowledge was partly conscious, as demonstrated by increasing accuracy when participants indicated some confidence in their decision. However, the process was also partly unconscious in that there was no significant relationship between confidence and accuracy, thus satisfying the zero-correlation criterion, which holds that a lack of correlation between confidence and accuracy is indicative of implicit knowledge (Dienes & Scott, 2005). Supporting the interpretation of the presence of implicit knowledge, an analysis of the data from the source attributions indicated that participants performed at above chance levels when making judgements based on intuition, which has also been argued to be reflective of unconscious judgement knowledge (Dienes, 2008). In summary, traditional analyses of the subjective measures of awareness demonstrate that the incidental exposure in the present study resulted in both conscious and unconscious knowledge on the part of the participants.

Assuming that these results do in fact reflect implicit knowledge, this finding runs partly counter to the results from Robinson (2002, 2005) and Grey et al. (2014), where the detected learning effects were found to be largely driven by explicit knowledge. This divergence from previous findings could be attributed to the smaller amount of input to which the participants were exposed in the training phase than was the case in other studies. In this study, participants received 72 exposures to each of the two case markers in the training set, or a total of 144 exposures. By contrast, in Robinson's (2002, 2005) study, participants received 450 total exposures, 150 for each of the three case endings. It seems plausible that such a difference in the total amount of input could result in increased opportunities to notice the morphological forms during the training phase, and to develop greater awareness of the underlying morphological rule system.

In Grey et al. (2014), however, the training phase consisted of only 128 exposures presented auditorily, which is comparable to the type and amount of exposure in the present study. It is worth noting, however, that the stimulus material in Grey et al.'s study comprised English words with Japanese case markers added to the end (*Vet-ga injection-o gave*). A number of studies have demonstrated that enhancing the target structure in the input, for example by underlining the target form within a reading text or presenting the target structures in bold, can promote the noticing, and subsequent acquisition of the target linguistic structure (Alanen, 1995; Leow, 2001). Given the results of these studies, and the argument that salience is a contributing factor to whether particular linguistic structures are noticed in the input (N. Ellis, 2006a; Peters, 1985; Schmidt, 2001; Slobin, 1985), it follows that attaching the morphological marker to the end of the English word could have increased the salience of the case markings, thus leading to heightened levels of awareness throughout the exposure phase.

However, at this point, it is worth offering an alternative interpretation of the patterns observed above by questioning the assumption that the results of the confidence ratings and source attributions are necessarily indicative of unconscious knowledge. While the data from the confidence ratings met the zero-correlation criterion, the lack of a relationship between confidence and accuracy could be a relic of the slight overall learning effect. This interpretation of the results reflects the view that it is probably not possible to develop grammatical, native-speaker-like intuition as a result of a single laboratory training session (Leow & Hama, 2013), and a recent suggestion by Serafini (2013) is that source attribution data could be interpreted differently. According to Serafini, source attributions can reflect 'degrees of explicitness', with above-chance performance for intuition responses indicating low-levels of awareness or knowledge that has been noticed, while above-chance responses for rule responses reflect awareness at a higher level of understanding. This interpretation runs counter to the traditional interpretation regarding the phenomenology of intuition reflecting native-speaker-like implicit knowledge (Dienes, 2008; see also Rebuschat & Williams, 2012). Serafini's (2013) interpretation would appear to sit well with the results of the present study, particularly regarding the reports of noticing in the retrospective verbal reports. Taken together, the results from both the subjective measures and the retrospective verbal reports in the present study would point towards low levels of awareness amongst the participants in this study. Subsequently, the results here would fit well with those of previous research (Brooks & Kempe, 2013; Grey et al., 2014), which revealed an important role of awareness in the acquisition of inflectional morphology.

6.1.4. Summary and suggestions for future research

The results of Experiment 1 suggest that participants developed some knowledge of the case-marking system as a result of incidental exposure, and in the absence of verbalisable rule knowledge. This knowledge, however, appears to be partial in that participants' performances were only significant for test items targeting the accusative, but not the nominative, case marker. In addition, although participants were not able to verbalise the rules for the case marking system, data from the subjective measures of awareness indicated that the participants developed low levels of awareness as a result of the training phase of the experiment. Thus, in summary, the results suggest that inflectional morphology can be learned, at least partially, as a result of incidental exposure, but that awareness might play an important role in this process.

Despite the apparent success of Experiment 1, the results thereof raise several questions to be addressed by subsequent experiments. Perhaps most crucially, due to the relatively small learning effect demonstrated here (approximately 56%), it would be prudent to establish that the overall learning effect, as well as the significant performance for accusative and Type 1 syntactic items, was indeed present, and was not the result of a Type 1 error (Brown, 1988, 2015). To address this question, a replication of Experiment 1 is warranted. As well as confirming the results of Experiment 1, a follow-up study to this experiment could also investigate a number of additional areas of interest that were raised by Experiment 1. For example, it was hypothesised that the inclusion of repetition in the training phase of Experiment 1 helped to lead to the learning effect demonstrated in this study. It would thus stand to reason that maintaining or increasing the amount of repetition in the study that follows would also be conducive to learning.

Another area worth investigating might be in relation to comments about the "strangeness" of the O-V-S word order, and that fact that this strangeness appears to have aided learning within Experiment 1. Future research could capitalise on this finding by modifying the training tasks in a manner that makes particular elements more salient. For example, when we consider that the participants across all three experiments commented on the changing word order, one manipulation that might lead to more noticing would be to modify the syntactic patterns so that foreign nouns have fixed positions within the sentence for one of the case markers (such as instrumental), but not for the other cases (nominative and objective). Thus, this contrast, much like the contrast between the S-V-O and O-V-S word orders, might lead to increased noticing, and possibly to hypothesis formation during the training phase of the experiment.

Yet another unexplored variable is in relation to the total amount of exposure participants received during the training phase of the experiment. Recall that the total amount of exposure was kept constant at 144 sentences across Pilot Studies 1 and 2, as well as for Experiment 1. If we accept the truism that a sufficient amount of input is necessary for adult SLA to occur, then one possible explanation for the lack of learning in Pilot Studies 1 and 2, and the need for repetition in Experiment 1, is that sufficient exposure was not provided in the initial pilot studies for learning to take place incidentally. In addition, as discussed in Section 6.1.3 above, it has been questioned whether participants can develop native-speaker-like intuition as a result of the minimal amount of exposure that is typical of laboratory-based SLA research. The logical extension of this line of argument is that increasing the amount of exposure should, in theory, lead to more learning, and should also increase the possibility that learning might take place incidentally. Future experiments could attempt to capitalise on this observation by manipulating the experimental procedure to maximise the total amount of exposure within the time constraints of a single laboratory training session.

6.2. Experiment 2

As noted in the section above, the purpose of Experiment 2 was to replicate and extend Experiment 1. Replication, simply defined, is the reproduction of a study in order to come to a fuller understanding of, or to verify, the findings of the original study (Abbuhl, 2011; Earp & Trafimow, 2015; Gass & Mackey, 2007; Mackey, 2012; Mackey & Gass, 2005; Perry, 2011; Polio & Gass, 1997). Replications can be broadly categorised as exact, approximate, or conceptual (Polio, 2012; Porte, 2012). Exact replications copy the original experiment exactly; in other words, with exactly the same population, design, materials, and so on. Approximate, or partial, replications copy the original study, but change a non-major variable, such as changing the population of the study. This allows for the results to be immediately comparable with the prior study, but examines whether the results can be generalised; for example, to a new population. Conceptual replications address the same problem as the original study, but typically employ a different research design to address the problem (Porte, 2012).[21] Experiment 2, then, represents an approximate replication of Experiment 1 (see Table 6.8 below for a summary of the changes between these two experiments).

Although researchers have historically been drawn to the "glamour" inherent in conducting original research using novel research designs (Earp & Trafimow, 2015; Valdman, 1993), there are a number of reasons that replication is important in scientific research. Firstly, as suggested in Section 6.1.5, above, replications can support the results

[21] It should be noted, however, that the exact, approximate, and conceptual categories have been used inconsistently in the social sciences (see Polio, 2012, for a discussion), and that many so-called "replication studies" fall into a grey area in which it is not clear whether they represent a follow-up study or an examination of the generalisability of the results of the original study (Porte, 2012).

from the original study not being due to Type 1 or Type 2 errors; for example as a result of unrepresentative sampling (Mackey, 2012; Nassaji, 2012; Plonsky, 2012). Furthermore, it has been demonstrated that, even in the absence of Type 1 or Type 2 errors, individuals within a particular population, as well as individual participants when tested at different times, differ in ways that can impact on the results of a study (Bodenhausen, 1990). In this sense, replication is important not only because it increases confidence that a Type 1 or a Type 2 error has not occurred, but also because it allows for a more robust and accurate estimation of the true effect size of a particular treatment (Plonsky, 2012). In summary, replication studies are essential for confirming the results of research, to the point that some do not consider a study is complete until it has been successfully replicated (Bauernfeind, 1968[22]; Dienes, 2008; Engel & Schutt, 2010; Muma, 1993).

To return to the present experiment: As noted above, the purpose of Experiment 2 was to replicate and extend the findings of Experiment 1. Such replication is deemed necessary due to the small learning effect (approximately 56% accuracy in the GJT) demonstrated in Experiment 1, and the fact that a significant learning effect was not found in the previous two pilot studies. In short, if the results from Experiment 1 could be reproduced in an additional study, it could then be said that the results of Experiment 1 were 'real' and not due to statistical error with more confidence.

As in Pilot Studies 1 and 2, and Experiment 1, Experiment 2 set out to address RQ 1 and RQ 2 in order to investigate the degree to which Czech morphology, specifically the nominative marker –*a*, the accusative marker –*u*, and the instrumental marker –ou, can be acquired under incidental learning conditions. There were several differences in the

[22] Note for SLA researchers: Bauernfeind (1968) highlighted practices in the natural sciences, in which experiments are often repeated 10 - 20 times (!) prior to initial submission for publication.

experimental procedure and in the stimulus materials between Experiment 1 and Experiment 2. These changes are presented in Table 6.8 below. In addition, these changes, and the rationale behind these changes, are outlined in detail in the sections that follow.

Table 6.8.
Description of Changes to Training and Testing Phases between Experiments 1 and 2

	Training Phase						Testing Phase
	Exemplars	Repetition	Total Exposure	Case markers	Syntactic Patterns	Training Task(s)	Testing Instruments
Experiment 1	48	3x	144 (auditory only)	Nominative Accusative	2	1. Picture-matching, 2. Elicited imitations (repeat sentence), 3. Elicited imitations (repeat foreign word in insolation)	1. 48-item auditory GJT (48 transfer items) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire
Experiment 2	60	4x	240 (auditory only)	Nominative Accusative Instrumental	2	1. Picture-matching, 2. Elicited imitations (repeat sentence), 3. Elicited imitations (repeat foreign word in insolation)	1. 48-item auditory GJT (48 transfer items) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire

6.2.1. Methods

6.2.1.1. Participants

Forty-six undergraduate native-speakers of English at a university in the United Kingdom volunteered to take part in this study. Six of these volunteers were excluded on the grounds of majoring in linguistics, or because they had a background in a language with a rich morphological system. The remaining 40 participants (24 female, 16 male) were randomly assigned to an experimental group (n = 20; 11 female, 9 male) or a control group (n = 20; 13 female, 7 male). None of the participants selected for this study majored in linguistics or foreign languages. The ages of the participants ranged from 18 to 27 years (M = 19.70, SD = 2.14).

6.2.1.2. Research Ethics

This experiment was covered under the exception granted to previous experiments in this thesis by the research support office of Lancaster University. As with the previous experiments, informed consent was granted by each participant prior to beginning the experiment. Furthermore, as was the case with previous experiments, it was stressed that participation in this study was voluntary and that participants could withdraw at any point.

6.2.1.3. Stimulus Material

The first modification to Experiment 2 was with regard to the training and testing sets. Firstly, the number of morphological markers was increased from two (nominative, accusative) to three (nominative, accusative, instrumental) in both the training and in the testing sets. These morphological markers were identical to those used in Pilot Study 2, examples of which can be seen in Table 6.9 below.

Table 6.9.
Descriptions and Examples of the Three Morphological Categories

Morphological Category	Syntactic category in English	Example
Nominative	Subject	*The britva cut David's face at the sink last night.*
Accusative	Direct object	*Peter used a britvu in the bathroom today.*
Instrumental	Adverbial (meaning "to do/make something with an X")	*Anne cut her leg with a britvou in the morning.*

It should be noted here that the syntactic patterns were not modified between Experiment 1 and Experiment 2 (see Table 6.10 below). Using the same syntactic patterns would entail that the instrumental marker would always appear at the end of the sentence. This is in contrast to the nominative and accusative markers, for which sentence position would not be a reliable indicator of their function. It was hypothesised that the fixed sentence position of the instrumental marker would result in an increased salience of the instrumental marker, and thus increased noticing during the training phase of the experiment. Given this fixed position, it was also hypothesised that participants would develop verbalisable, explicit knowledge of this case marker; specifically, that participants would be able to identify that the *-ou* marker always occurred in conjunction with the prepositional phrase at the end of the sentence. In other words, it was believed that the participants would be able to 'crack the code' and correctly identify the *-ou* marker and its function within the sentence.

Table 6.10.

Templates, Sample Sentences, and Frequencies for the Two Syntactic Patterns included in the Training and Testing sets for Pilot Study 2

Pattern	*Template*	*Frequency in training set*
Pattern 1	$[[AP]_{TEMP} > [NP]_{OBJ} > [VP] > [NP]_{SUBJ} > [PP]]$	
Nominative	*Last summer the grass ate the koza in the field.*	(24)
Accusative	*Last month the kasu opened Patrick with the key.*	(24)
Instrumental	*Some time ago John scared the child with a zrudou.*	(24)
Pattern 2	$[[AP]_{TEMP} > [NP]_{SUBJ} > [VP] > [NP]_{OBJ} > [PP]]$	
Nominative	*Last year the prodejna shipped goods to the shoppers.*	(24)
Accusative	*All week the builder took his vrtacku to work.*	(24)
Instrumental	*Today the wooden board cut he with a pilkou.*	(24)

An additional change between Experiment 1 and Experiment 2 was that the training set was reduced from 24 nouns to include only 20 nouns.[23] These 20 nouns (which are highlighted in Appendix C) were repeated four times in nominative, accusative, and instrumental cases across four training blocks, resulting in a total of 240 total stimulus sentences (80 for each of the three case markers). This represented an increase in both the number of times the individual exemplars were repeated in the training set and an increase in the total amount of exposure during the training set. As the repetition of exemplars was implicated in the learning effect of Experiment 1, it was hypothesised that increasing the number of repetitions of individual exemplars, as well as an increase in the total amount of exposure, would further facilitate the acquisition of the target morphological markers.

6.2.1.4. Procedure

As was the case with Experiment 1, Experiment 2 took place in a private office on campus. The training and testing phases of the experiment were delivered via a Dell Thinkpad T410 laptop computer, using the stimulus presentation software Superlab 4.5

[23] The decision to reduce the total number of exemplars from 24 in Experiment 1 to 20 in Experiment 2 was based on issues of practicality, such as maintaining the total amount of exposure that was feasible in the time allotted to each participant, allowing the exemplars to be neatly divided into manageable training blocks, and so forth.

(Cedrus Corp, San Pedro, CA), together with a Cedrus model RB-834 response pad. All audio for the experiment was played through a set of Audio-Technica headphones with noise-cancellation functionality. Following the testing phase, participants completed a short debriefing questionnaire, followed by an oral interview. All participants in this study, in both the experimental and the control groups, gave their informed consent prior to the outset of the experiment.

In summary, the procedure for Experiment 2 was identical to that for Experiment 1, as can be seen in Figure 6D below.

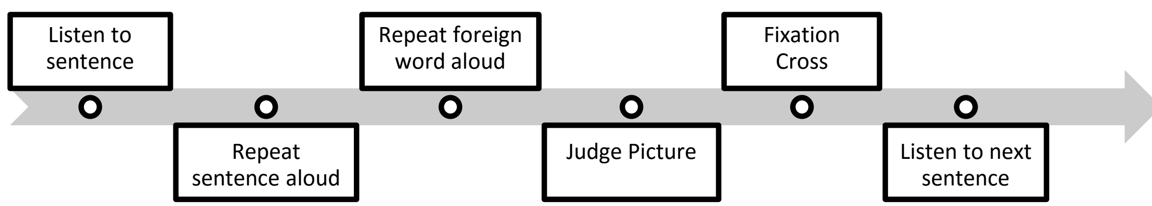


Figure 6D. *Training procedure in Experiment 2.*

At no point during the training phase were participants informed that the foreign nouns were inflected for case, nor that they would be tested afterwards. Participants listened to all 240 sentences without a break. These sentences were presented in a different, randomised order for each participant. The entire training phase took about 45 minutes to complete.

6.2.1.5. Statistical Analyses

As in Pilot Studies 1 and 2 and Experiment 1, performance in the grammaticality judgement task in Experiment 2 was analysed using mean accuracy rates and mixed-effects models.

6.2.2. Results

6.2.2.1 .Retrospective verbal reports

An analysis of the retrospective verbal reports indicated that all participants (20 / 20) reported noticing the morphological inflections at the end of the foreign words during the training phase of the experiment. As in Experiment 1, there were comments such as "the foreign words always ended in vowels", "the endings of the words changed", or comments that specifically identified the foreign markers. Such comments were taken as evidence that the learners had attended to, and had subsequently become aware of, the surface features of the morphological markers. When specifically asked about the endings of the foreign words in particular, the majority (14/20) of the participants were able to recall all three of the morphological markers (-*a*, -*u*, -*ou*), with the remaining participants (6/20) being able to recall two of the three morphological markers. Interestingly, all six of these participants were able to recall the nominative marker (-*a*); however, only three participants recalled the accusative marker (-*u*), and three participants the instrumental marker (-*ou*).

Despite these reports of noticing the changing morphological markers, none of the participants were able to demonstrate an understanding of the underlying rule system by verbalising the rules governing the use of the inflections. When prompted to guess, ten participants stated that the affixes might have represented a noun class, i.e., gender. Five participants mentioned that they thought that the inflection might be connected with the position of the foreign word in the sentence. Three of these participants (3/5) could not explain the nature of this connection. Two participants, however, stated that they initially believed that the -*ou* marker was occurring at the end of the sentence, but later abandoned this theory because they felt it was also sometimes occurring in other parts of the sentence. The comments of these two participants suggest that it is possible that the perceptual salience of the *-u* and *-ou* markers might have led participants to conflate these two case endings. In addition, concerning the construct validity of the incidental training conditions, these comments clearly point toward conscious rule-search behaviour and hypothesis formation during the training phase of the experiment.

At the very end of the experiment, when the rules were explained to the participants, none of the participants claimed that he or she had thought of the rules related to the nominative or accusative items at any point during the experiment. Five of the participants replied with comments such as "I thought so" or "that makes sense now" when the rule for the instrumental marker was explained to them. In summary, as in Experiment 1, the analysis of the retrospective verbal reports suggests that participants exhibited awareness at the level of noticing, but still below the threshold of verbalisation (Schmidt, 1990) for the nominative and accusative markers. However, with regard to the instrumental marker, it appears that some of the participants had developed awareness at the level of understanding; however, due to being uncertain or unsure, they did not report this data freely until the rule was explained to them.

6.2.2.2. Picture-matching task

As was the case with Pilot Studies 1 and 2, and Experiment 1, the participants were able to complete the picture matching task in the training phase with a near perfect level of accuracy (M = 98.76, SD =0.77). This again suggests that the training task was successful in orienting the participants' attention towards the meaning of the sentences in the training phase of the experiment.

6.2.2.3. GJT

As in Experiment 1, mean accuracy rates were generated to examine the performance of the experimental and control groups in the GJT. With regard to the overall performance, the accuracy of the experimental group (*M* = 54.06%, *SD* = 7.53%) appeared to be slightly higher than that of the control group (*M* = 47.40%, *SD* = 9.28%). Furthermore, as can be seen in Table 6.11 below, breakdowns of performance across the three case markers in this study indicate that experimental group was more accurate in accusative case items, than they were with nominative and instrumental items. Further breakdowns revealed that the experimental group performed with higher degrees of accuracy on ungrammatical versus grammatical items, as well as on items that followed an O-V-S word order, as opposed to an S-V-O word order. This is contrast to the performance of the control group, which appeared to perform similarly across the three different case markers, grammatical versus ungrammatical test items, and across the different syntactic patterns. At a cursory glance, these results appear very similar to the results from Experiment 1.

Table 6.11.
Results of Experiment 2: Descriptive Statistics (%) for Experimental and Control Groups across Grammatical, Ungrammatical, and Type 1 and Type 2 Syntactic Patterns

Item Type	Descriptive Statistics (%)		
Group	M	SD	SE
Grammatical			
Experimental	51.87	11.82	2.64
Control	47.61	12.79	2.86
Ungrammatical			
Experimental	56.25	14.15	3.16
Control	47.19	13.40	3.00
O-V-S			
Experimental	57.19	10.46	2.34
Control	46.25	10.20	2.28
S-V-O			
Experimental	50.83	9.42	2.11
Control	48.75	12.54	2.81
Nominative			
Experimental	48.44	13.43	3.00
Control	47.19	11.38	2.54
Accusative			
Experimental	60.31	9.13	2.04
Control	48.44	13.29	2.97
Instrumental			
Experimental	53.13	10.04	2.24
Control	47.19	13.37	2.99

In order to determine whether these apparent differences in performance between the control and experimental group were statistically significant, a logit mixed-effects regression model (Jaeger, 2008) was carried out using R (R Core Team, 2014) and lme4 (Bates et al., 2014). This model was constructed following exactly the same procedure used in Experiment 1. Firstly, analyses were carried out to determine the fixed effects that had a significant relationship with the dependent variable, namely performance in the GJT. To this end, a series of logit mixed-effects models were built. For these analyses, a null model was constructed, which analysed performance in the GJT as a binary outcome in terms of subject and item as crossed random intercepts. Additional models were then built for each fixed effect. Each of these models included a variable, such as case, as a fixed effect as well as a random slope (Barr et al., 2013). As was done in the previous analyses in this thesis, these models were then compared individually against the null model using the $\chi 2$ statistic. As noted previously, if the model that included the fixed effect was significant against the null model, this result was interpreted as indicating that the fixed effect in question had a significant relationship with the dependent variable, and should be included in subsequent analyses. If the result was nonsignificant, this was interpreted as no significant relationship being present, and that this fixed effect could be excluded from the models that followed. The results of these comparisons can be seen in Table 6.12 below.

Table 6.12.
Summary of Likelihood Ratio Tests on Predictor Variables in GJT

Predictor Variable	χ^2	df	p	C
Group	199.24	3	<.001***	.78
Case	19.20	4	<.001***	.78
Syntax	18.33	6	.48	.77
Grammaticality	5.26	7	.51	.77

As can be seen in Table 6.12 above, the results of these comparisons indicated that the fixed effects of case (nominative versus accusative versus instrumental) and group (experimental versus control) showed a significant relationship with performance in the GJT. Therefore, a new model that included both of these variables as fixed effects was constructed. In addition, following the maximal model as suggested by Barr et al. (2012), the fixed effect of case was included as a random slope by participant, and group as a random slope by item. A maximum likelihood ratio test indicated that this model was significant against the null model, χ^2 (12) = 205.89, $p < .001$, C=.78. Further comparisons indicated that the inclusion of both case and awareness as fixed effects was warranted in the model, and the removal of either fixed effect did not lead to an improved model fit. Multicollinearity within this model was formally assessed using Cohen's kappa. This resulted in a kappa value of 6.88 which, according to Baayen's (2008) guidelines, indicates an acceptable level of collinearity between the fixed effects in the model. Thus, this model of the data from the GJT is henceforth referred to as the best-fit model, the results of which are presented in Table 6.13 below.

Table 6.13.
Results for the Best-fit Logit Mixed-effects Model Examining Performance in the GJT

					Random Effects	
		Fixed Effects			By subject	By items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	-0.13	.18	-.77	.44	.33	2.13
Group	0.55	.16	3.38	<.001***	–	1.73
Case	-.28	.23	-1.24	.21	.18	–
Group: Case	-.63	.23	-2.74	<.01**		

Note. factors were coded using contrast coding, as follows: Case (−1= Nominative, 0= Accusative, 1 = Instrumental), Group (-.5 control, .5=experimental), formula: "correct~case*group + (case|subject) + (group|item), family=binomial, control= glmerControl (optimiser = "bobyqa")

As can be seen in the table directly above, the fixed effect of group was significant within the model. However, unlike Experiment 1, the fixed effect of case did not reach significance as a main effect. It is important to emphasise that "a main effect need not be significant if it is involved in interactions which are significant" (Baayen, 2008, p. 166). In this regard, the model revealed that the fixed effect of case interacted significantly with the fixed effect of group. To analyse these results further, post-hoc Tukey tests were carried out to compare the performance of the experimental group and the control group across the three different case markers. These analyses revealed that the performance of the experimental group was significant against the performance of the experimental group for accusative items. However, this significant performance did not extend to the other two case markers. A summary of the results of the Tukey tests can be seen in Table 6.14 below.

Table 6.14.
Results of Experiment 2: Descriptive Statistics (%) and Results of Tukey Post-hoc Tests for Experimental and Control Groups across Nominative, Accusative, and Instrumental Case Markers

Item Type	*Descriptive Statistics (%)*			Results of Independent Samples T-tests				
Group	*M*	*SD*	*SE*	*B*	*SE*	*z*	*p*	*d*
Nominative								
Experimental	48.44	13.43	3.00	-0.08	0.16	-0.49	0.99	0.10
Control	47.19	11.38	2.54					
Accusative								
Experimental	60.31	9.13	2.04	0.49	0.16	2.989	0.03	1.04
Control	48.44	13.29	2.97					
Instrumental								
Experimental	53.13	10.04	2.24	0.35	0.16	2.115	0.27	0.50
Control	47.19	13.37	2.99					

6.2.2.4. Subjective Measures of Awareness

Source attributions. As indicated in Table 6.15 below, an analysis of the source attributions revealed that participants relied largely on their intuition rather than on explicit rule knowledge in their judgements. This finding mirrors the results of Experiment 1, in which participants’ source attributions also indicated that they based most of their decisions on their intuition. However, unlike Experiment 1, the performance of the participants was not significant from chance (50%) levels when basing their decisions on intuition, or on any of the other options available. This indicates that the guessing criterion for implicit knowledge was not met by the data here (this point is discussed further in the following table, Table 6.15 below).

Table 6.15.
Accuracy (%) and Number of Responses for Experimental Group Participants across Confidence Ratings and Source Attributions (OVERALL)

	Accuracy (%)	Number	*p*-value
Confidence Rating			
No confidence	52.2	182	.56
Somewhat confident	53.9	480	.09
Very confident	54.7	260	.14
Absolutely certain	63.4	38	.09
	Accuracy (%)	Number	*p*-value
Source Attribution			
Guess	50.9	221	.78
Intuition	54.2	432	.08
Memory	56.2	144	.14
Rule	55.4	163	.16

Confidence ratings. As shown in the table above, participants' performances were nonsignificant against chance levels for all possible ratings. As in Experiment 1, an additional logit mixed-effects model with crossed random effects of subjects and items (Jaeger, 2008) was constructed in order to analyse if there was any underlying interaction between confidence and accuracy. In this model, accuracy was modelled as a binary outcome, and confidence level was analysed as a fixed effect, along with crossed random intercepts for subjects and items. A likelihood ratio test revealed that the levels of confidence and accuracy were not significantly related $\chi 2$ (1) = 0.92, p =.34, C=.67. This result satisfies the requirements for the *zero correlation criterion* (Dienes & Scott, 2005), and suggests that the knowledge developed by the participants was at least partly implicit in nature.

With regard to the results for the subjective measures of awareness presented above, it is worth returning to a previous point concerning the results of Pilot Studies 1 and 2. Subjective measures of awareness are typically not analysed in studies that report nonsignificant results. The rationale for this is that a nonsignificant finding indicates that no knowledge, either implicit or explicit, has been developed as a result of the exposure phase. In the case of the present study, analyses revealed that performance was significant against chance-level performance for accusative items, but not for nominative and instrumental items. This would suggest that the presence of the nominative and instrumental items might be skewing the subjective measures data. In other words, isolating the accusative items within the subjective measures of awareness data might provide a more refined picture with regard to the type of knowledge acquired by the participants in this study.

Source attributions: Accusative items only. As can be seen in Table 6.16 below, an analysis of the source attributions for accusative test items reveals several significant findings. Firstly, as in the overall analysis presented in Table 14 above, participants here also tended to rely on their intuition when making their judgements. In contrast to previous findings, their performance was significant compared to chance levels when attributing their judgements to the categories of *intuition* and *memory*. Performance when attributing their decisions to *guessing* or *rule* was nonsignificant against chance-level performance. These data indicate that the guessing criterion for implicit knowledge was not met by the accusative data here.

Table 6.16.
Accuracy (%) and Number of Responses of All Experimental Group Participants for Accusative Items across Confidence Ratings and Source Attributions

	Accuracy (%)	Number	*p*-value
Confidence Rating			
No confidence	53.5	69	.56
Somewhat confident	58.1	162	.03
Very confident	68.5	73	.001
Absolutely certain	75.0	16	.041
	Accuracy (%)	Number	*p*-value
Source Attribution			
Guess	50.0	81	1.0
Intuition	62.3	137	.003
Memory	70.8	48	.003
Rule	61.1	54	.10

Confidence ratings: Accusative items only[24]. As shown in the table above, participants' performances were significant against chance levels when indicating that they were *somewhat confident*, *very confident*, or *absolutely certain* about their decisions. However, their performance when indicating *no confidence* was not significant in contrast to chance levels. To examine whether there was a relationship between confidence and accuracy, an additional logit-mixed effects model (identical to the analysis performed on the overall data set above) was constructed. A likelihood ratio test revealed that the levels of confidence and accuracy were not significantly related: $\chi2$ (1) = 2.31, *p* =.12, C=.55. This would suggest that the accusative data here satisfy the requirements for the *zero correlation criterion* (Dienes & Scott, 2005), and that the knowledge developed by the participants was at least partly implicit in nature.

[24] Analyses of the subjective measures for nominative items and instrumental items are not included here, as performance on these items was not statistically significant compared to chance levels (50%).

This result is, however, surprising, given that there appears, prima facie, to be a linear relationship between the accuracy and the level of confidence, as reported in Table 6.15 above. To investigate this further, the data related to the accusative confidence ratings were re-grouped into two categories, a *less confident* category and a *more confident* category. The less confident category included responses whereby the participants indicated *no confidence* or *somewhat confident*, and the more confident category included *very confident* and *absolutely certain* ratings. Another logit mixed-effects model, as described above, was then constructed to compare the relationship between accuracy and confidence across these two categories. Here, a likelihood ratio test revealed a significant relationship between confidence and accuracy; $\chi^2(3) = 8.00$, $p = .04$, $C = .66$). These results indicate that participants' decisions were more accurate when they indicated higher levels of confidence (such as *very confident* or *absolutely certain*) than when they indicated lower levels of confidence (*no confidence* or *somewhat confident*), thus providing further evidence that the participants' knowledge of accusative items was more explicit in nature.[25]

6.2.2. Interim Discussion of Experiment 2

The results of Experiment 2 support the findings of Experiment 1; the data here indicate that participants can develop some knowledge of L2 case markings after a minimal amount of exposure, without feedback, and as a result of incidental exposure. With regard to the relationship between learning and awareness in this experiment, an analysis of the retrospective verbal reports indicates that this learning effect was realised in the absence of verbalisable rule knowledge. However, participants demonstrated low

[25] It is acknowledge here that running multiple analyses on the same dataset increases the likelihood of a "false positive" result, namely a Type 1 statistical error (Brown, 2015).

levels of awareness of morphological markers, in that they became aware of the surface features of these forms during the training phase of the experiment. Although traditional analyses of the subjective measures of awareness (guessing criterion, zero-correlation criterion) suggest that the exposure phase had resulted in some implicit knowledge of L2 morphology amongst the experimental subjects, finer-grained sub-analyses indicated that the overall learning effect was driven by conscious knowledge of the accusative marker.

Research Question 1: To what extent can L2 case marking be acquired under incidental learning conditions?

One finding of particular interest is the similarity between the results of Experiment 1 and Experiment 2. Specifically, in both experiments, the experimental groups' performance was significant for accusative, but not for nominative, case items. In the discussion of Experiment 1, it was hypothesised that the differential performance for accusative versus nominative items could be related to the phonotactic qualities of the respective case markers, specifically because the nominative marker, *–a* (/a/), was probably perceived by English speakers as more natural in a word-final position than was the phoneme used to realise the accusative marker *–u* (/u/), given the phonotactic rules of the English language. As a result, the "oddity" of the accusative marker could have made this marker more salient, resulting in greater attention being paid to, and awareness of, these items during the training phase of the experiment. While this explanation appears plausible, it is speculative at this point, as there is no direct evidence to support this interpretation. Future experiments might test this hypothesis by rotating the case endings between the nominative and accusative markers. This manipulation would provide direct support for whether the learning effect for accusative items is directly related to the phonotactic qualities of the accusative marker, or whether it is due to another, unaccounted for, variable.

Another explanation for why participants in both Experiments 1 and 2 developed some knowledge of the form-meaning mapping of accusative markers, but not of nominative markers, might lie in the nature of learning case-marking systems in general, and the universal constraints that bind this process. It has been well documented in both the L1 and in the L2 literature that learners incorporate different grammatical features of case-marking systems at different points in their development (Bardovi-Harlig, 2000; Eisenbeiss, Narasimhan, & Voeikova, 2008).[26] In Slavic languages, such as Czech – the language on which the current artificial system is based - the nominative form is the default and is the first one incorporated into the child's interlanguage as part of the L1 acquisition process (Dabrowska & Szczerbinski, 2006; Gagarina & Voeikova, 2009)[27]. However, within formal linguistics, it has been argued that the accusative case is the default in English (Emonds, 1985; Schütze, 2001; see Eisenbeiss, Bartke, & Clahsen, 2006 for an overview).[28] A possible explanation for the fact that the population of this study, L1 speakers of English, appeared to acquire the accusative marker could be attributed to an L1 transfer effect. In further support of this point, the nature of the L1 has been implicated in the degree that adults process, and subsequently acquire inflectional

---

[26] This is in no way meant to suggest equivalence between L1 and L2 acquisition. The L1 literature with regard to the acquisition of case marking in naturalistic settings is particularly rich. It is not unreasonable to presume that this body of literature could provide some insights that are relevant to the discussion of L2 learning.
[27] The default nominative form has been argued to be a result of frequency in the input. For example, Dabrowska and Szczerbinski (2006) pointed out that nominative forms have been shown to account for more than 40 per cent of the inflected forms in child-directed speech.
[28] Examples of this can be seen in the use of nominative versus accusative pronouns in Czech versus English. In Czech, the answer to the question "Who is it?" would be "To jsem Ja!" literally, 'It am I'. In English, a more natural answer would incorporate the accusative pronoun "me": for example, "It's me". For a more detailed discussion, see Schütze (2001).

morphology in an L2 (Brooks & Kempe, 2013; Kempe & MacWhinney, 1998; see Gor, 2010 for an overview).

Another possible explanation might be in relation to the size of the learning effect exhibited in Experiments 1 and 2, the acquisition of case-marking systems is typically operationalised in terms of a child’s ability to produce utterances that demonstrate a contrast between two different forms, such as nominative versus accusative (Penke, 2012). In Slavic languages, such as Czech, the initial contrasts that children exhibit are nominative versus vocative forms, and/or nominative versus accusative forms (for an overview, see Gagarina & Voeikova, 2009). Children's initial L1 development of case-marking systems is divided into two stages, a pre-morphological stage and a proto-morphological stage (Gagarina & Voeikova, 2009; Dressler, 2012; Slobin, 1995). Within the pre-morphological stage, correct instances of inflectional morphology can be found in the child’s output, but these correct forms represent instances of rote-learned grammatical utterances and formulaic speech, rather than a contrast between the form-meaning mappings of different morphological markers. Within the proto-morphological stage, children begin to produce utterances that demonstrate a contrast between different case endings. This stage is characterised by the creative construction of morphological forms, although these are not error free.

When we compare how the acquisition of case-marking is operationalised in the L1 literature with the results of the present study, it would appear that the participants here did not acquire the ability to differentiate between the nominative and accusative markers, given the null results for the nominative marker and the slight learning effect demonstrated for the accusative marker. However, when taking into account the slight learning effect demonstrated in the GJT for accusative items, one might argue that the results here represent the very earliest stages of the acquisition process. Given the very limited amount of exposure in the training phase of the present study, this might seem to be a plausible explanation. Such an explanation appears speculative, however, given the lack of direct evidence that corresponds to the L1 stages of development outlined above. For instance, the pre-morphological stage, which represents the initial stages of learning in the L1 model, is characterised by rote-learned forms. However, the GJT in the present study consisted entirely of transfer items, or novel forms. Although great caution should be employed when comparing the stages of L1 acquisition with the acquisition of an L2, it could be said that the use of transfer items might not be a valid measure of the learning that did occur in the present study (Clahsen, Eisenbeiss, & Penke, 1996; see also Hulstijn, 2013 for a discussion of transfer appropriate measures in SLA). Future experiments could investigate this issue by including a second GJT that contains only trained items; in other words, items that the participants have previously been exposed to during the training phase of an experiment.

To return to the safe ground of the L2 literature, there is also support for the claim that adult learners go through a number of developmental stages with regard to the attainment of morphological accuracy. For example, in what is known as the one-to-one principle (Andersen, 1984), it has been documented that L2 learners first learn a single form-function mapping for an individual morphological form. Learners then progress through a number of additional stages in which they begin to use the form in an increasing number of contexts and functions, before finally arriving at a full form-function mapping (Cadierno, 2000; Shirai, 2002, 2004). However, despite the fact that evidence for “developmental stages” might be interpreted as suggesting that morphological competence develops in a steady, systematic fashion, it should be pointed out that morphological development is, in fact, characterised by a non-linear path of development marked by both decreases and increases in accuracy (Bardovi-Harlig, 2000; Ortega, 2009), and furthermore, it develops late, if at all, in comparison to other linguistic areas. Evidence for the late development of L2 morphology can be seen in a frequently cited, longitudinal study by Klein and Perdue (1997). In this study, the researchers followed the time course of second language development for a group of 40 learners from a variety of linguistic backgrounds. After some time, the researchers noticed that all of the learners' interlanguages had stabilised in a relatively simple system on which these learners relied to express themselves. This simple system was referred to as the basic variety. The authors noted that many of these learners did not progress beyond the basic variety, and they hypothesised that it was the need to express more complex thoughts that was the driving force in the development of more complex language. In other words, in the absence of any need to create complex utterances to express meaning, learners in naturalistic settings will not develop complex interlanguage. Importantly, the basic variety includes no morphological markers, suggesting that learners are unlikely to develop morphological competence, unless there is a communicative need to do so. When we consider the present experiment, the fact that the meaning carried by the case markers was redundant would have likely impeded the learning of their form-meaning mapping, in that there was not an immediate communicative "need" for the participants to attend to the meaning carried by the inflectional forms.

With regard to the differential performance demonstrated for the different case markers, another interesting finding was the lack of a strong learning effect for instrumental case items. As noted previously, it was hypothesised that the fixed sentence-final position of the instrumental marker would facilitate the acquisition of this marker. This was not the case. One possible explanation for the low amount of learning demonstrated for instrumental items might be the phonological similarity between the instrumental (/oʊ/)and accusative (*/u/*) markers. Additional support for this explanation comes from the retrospective verbal reports, in which all participants reported noticing the nominative marker (-*a*), whereas a quarter of the participants (5/20) did not appear to have differentiated between the accusative and instrumental markers. Although this explanation would seem plausible given that phonological distinctiveness has been identified as a factor that impacts on the acquisition of morphology (Goldschneider & DeKeyser, 2001), a question worth asking is whether this also impacted on the performance of the experimental group for accusative items. It remains possible that the learning effect for accusative items was mitigated by the phonological similarities between the accusative and instrumental markers. Future research might investigate this issue by modifying the stimulus set to include more phonologically distinct morphological markers (Jackson, 2014). Such a modification might allow researchers to better account for the different performances on individual case markers that have been exhibited here.

Yet another interesting finding was that the increase in the total amount of exposure in Experiment 2 did not lead to an increase in learning when compared to Experiment 1. As noted previously, the training phase of Experiment 2 included 240 total exposures, whereas Experiment 1 contained only 144. It was hypothesised that this increase would lead to an increase in learning. However, this change did not result in a clear increase in performance. The simplest explanation for this is that the increase in total exposure was not sufficiently significant to impact demonstrably on learning. Given that SLA is generally held to be a slow, incremental process that takes place over hundreds, if not thousands of hours of learning, it would seem reasonable that the effects of an additional 15 minutes of exposure might not be immediately reflected in performance. Notwithstanding this explanation, there are a number of other factors that might have mitigated the benefits of the additional exposure. First, the present experiment investigated the learning of three morphological markers as opposed to the two markers in Experiment 1. This increase in complexity might have masked the benefits of the increased input. Another possibility lies in the U-shaped development, which has been argued to be characteristic of morphological development (Long, 2015; Ortega, 2009). Studies investigating the time course of learning (Karuza et al., 2014) have demonstrated that learners quickly become sensitive to the regularities of input. It seems possible that the increase in input led to diminishing returns on learning, in that the learners became used to, and subsequently paid less attention to, the input as the experiment progressed. It also seems possible that the present experiment might have included too much input, resulting in fatigue and lower rates of performance. Still another explanation might lie in the decreased type of variability of the Czech nouns included in the training set. Previous research in artificial grammar learning (Gómez, 2002) and adult SLA (Brooks et al., 2006) has indicated that exemplar variability impacts on the degree to which learners are able to generalise morphological markers. In particular, Brooks et al. (2006) indicated that learners who were exposed to fewer exemplars in the training set were less successful when generalising this knowledge to novel words.[29] As the testing phase of Experiment 2 contained only transfer items (items that measured the participants' ability to generalise their knowledge), the reduction of exemplars in the training phases from Experiment 1 (24) to Experiment 2 (20) might have influenced the amount of learning across these two experiments.

Research Question 2: What type of knowledge is acquired; implicit or explicit?

As with the results of Experiment 1, the results of the retrospective verbal reports indicated that none of the participants had become aware of the morphological rule system with regard to the nominative and accusative markers. By contrast, for the instrumental markers, the reports from some of the participants indicated that they did possess a higher level of awareness. It should be noted, however, that this higher level of awareness reported for the instrumental marker did not translate into a significant learning effect. Conversely, performance in the GJT was significant for accusative markers, for which no verbalisable rule was reported. As in Experiment 1, the lack of verbalisable knowledge in this experiment provides further support for the notion that subjects can develop knowledge of L2 morphology without being able to account for their performance verbally. It is worth reiterating that although none of the participants in this study were able to fully verbalise the underlying rule system, all of the participants reported noticing, or becoming aware of, the changing inflections during the training phase of the experiment. This report of noticing is indicative of a low level of awareness of the target linguistic forms (Schmidt, 1990), thus suggesting, as do the results of

[29] It is interesting that the effects of type variability in this study were mediated by scores on a culture fair test, suggesting that cognitive individual difference variables also play an important role in this process.

Experiment 1, as well as previous research in this area (Brooks & Kempe, 2013; Grey et al., 2014; Robinson, 2002, 2005), that attention and awareness might be closely tied to learning this area of grammar.

With regard to the subjective measures of awareness, initial analyses of the data in Experiment 2 indicated that the knowledge acquired of the morphological markers was implicit in nature. For example, the first mixed-effects model generated for the confidence ratings found no relationship between confidence and accuracy in the GJT. In other words, this initial analysis suggested that the zero-correlation criterion of implicit knowledge was met (Dienes, 2008), and that participants in this experiment were unaware of the knowledge they had gained as a result of incidental exposure. Nonetheless, it was questioned whether this result was necessarily reflective of implicit knowledge, or was simply a relic of the low levels of learning demonstrated in this study. A finer-grained analysis of the confidence ratings, which included only accusative items, pointed towards a different pattern in the data. The observation here was a clear relationship between confidence and accuracy. This observation also received

statistical support in that an additional logit mixed-effects model indicated a significant relationship between confidence and accuracy. This result points to the fact that participants' knowledge of the accusative marker was conscious in nature.

Supporting the interpretation of the presence of conscious knowledge, an analysis of the data from the source attributions indicated that participants' performance when attributing their decisions to chance was not significant. This means that the guessing criterion for implicit knowledge was not met (Dienes & Scott 2005). However, it should be noted that participants' performance across all source attributions was nonsignificant.

Again, this result is best seen as a reflection of the near chance-level performance of the experimental group in the GJT. As was done with the confidence ratings above, the source attribution data were reanalysed using only accusative-item data. The result of this analysis indicated that participants' performance was significant when decisions were attributed to intuition and memory, but not when attributing decisions to guessing or rule-knowledge. If we return to Serafini's (2013) interpretation of source attribution data, in which the options of *guessing* and *intuition* reflect low levels of awareness, and the choices of *memory* and *rule* indicate higher levels of awareness, then it would appear that the source attribution data point to a low-to-medium level of awareness among participants in this experiment. In other words, the data suggest that the participants developed awareness at the level of noticing, but below the level of understanding. This interpretation is also corroborated by the results of the retrospective verbal reports, in which participants also reported low levels of awareness, but were unable to verbalise the rules underlying the morphological system. In summary, the results of the measures of awareness in Experiment 2 provide little evidence that learning occurred in the absence of awareness. Instead, the learning that did take place appears to be marked by low levels of awareness in the form of noticing (Schmidt, 1990).

6.3. Future Directions

In summary, both Experiment 1 and Experiment 2 demonstrated that learners can acquire some knowledge of L2 inflectional morphology under incidental learning conditions following a minimal amount of exposure. In addition, the measures of awareness from both studies indicated that this knowledge was accompanied by, at a minimum, low levels of awareness. Despite the fact that both studies demonstrated similar results, it is important to note that the learning displayed in both experiments was minimal, suggesting that the knowledge developed by the learners of the underlying case-marking system was only partial in nature. An important question to be raised at this point is whether the instrument used to measure learning in Experiment 1 and Experiment 2 necessarily provides an accurate reflection of the quality and quantity of learning that might have taken place in the experiment. As noted previously, the GJT was chosen for these experiments as it is the instrument most commonly used in the AGL paradigm. However, the transfer appropriateness of this instrument is worth questioning (Hulstijn, 2013; Larsen-Freeman, 2013; Lightbrown, 2008). In other words, it has been argued that untimed GJTs, as used in the present study, are more likely to be accurate measures of the knowledge that develops as a result of explicit instruction, and may not provide an accurate measure of the type of knowledge resulting from incidental exposure (Norris & Ortega, 2000). This argument would appear to sit well with the results of Experiments 1 and 2, in which the learning that was displayed in the present experiment was accompanied by low levels of awareness. It remains possible, then, that the learners developed some knowledge of the case-marking system, which was not revealed due to the relative insensitivity of the instrument used in these studies.

It is also worth highlighting that much of the discussion surrounding the lack of a significant learning effect in Pilot Studies 1 and 2 and the learning that did occur in Experiments 1 and 2 has focused on issues related to attention, awareness, and processing. More specifically, the impact that the modifications between experiments, such as the addition of the elicited imitations, have been discussed with regard to the degree to which they resulted in increased attention to and/or processing of the target morphological forms during the training phase of the experiment. Again, this discussion has been speculative in that none of the previous experiments included any online measures of attention and/or processing during the training phase of the experiments. If attention to and the processing of the target grammatical features are driving the learning effect in these studies, then future experiments might benefit by employing alternate instruments that would allow for data to be gathered directly at the point of learning.

Within both cognitive psychology and SLA, researchers have turned to alternate measures of learning that incorporate “real-time” measures of processing and learning, such as reaction-time data. Such methodologies have proven fruitful in SLA in investigating the role of awareness in the incidental learning of form-meaning connections (Leung & Williams, 2011, 2012, 2014), word-order regularities (Williams & Rebuschat, 2013), and differential object marking (Andringa & Curcic, 2015). It seems that one avenue for future exploration might be to address the research questions of this thesis from a different methodological angle, specifically by modifying the research design to include an online measure of processing and learning. This might shed light on the underlying processes that might influence the incidental learning of an L2 case-marking system.

CHAPTER 7

EXPERIMENT 3

This chapter reports on Experiment 3, the final experiment of this thesis. Experiment 3 utilised a self-paced reading methodology to address the following research questions:

1) To what extent can L2 case markings be learned under incidental learning conditions?

2) What type of knowledge is acquired as a result of this exposure; implicit or explicit?

This chapter is organised as follows. It begins by providing a brief background to and rationale for Experiment 3, which is followed by a more detailed description of the population, methods, materials and experimental procedure. The results of this experiment are then presented, followed by a discussion. The chapter concludes with a summary of the findings, including a critical evaluation of the research design utilised, as well as recommendations for future experiments.

As noted above, Experiment 3 signals a departure from the AGL research methodology that was used in previous experiments in this thesis. Instead, Experiment 3 utilises a self-paced reading (SPR) task to address the research questions set out above. The rationale for using an SPR task is as follows. As discussed in the literature review, SPR tasks have been used effectively for a variety of purposes within both psycholinguistics and SLA (Jegerski, 2014; Juffs, 2001; Keating & Jegerski, 2015; Roberts, 2012a), and have been referred to as the "most fundamental experimental measure" in psycholinguistic research investigating L1 and/or L2 sentence processing (Jegerski, 2014). Furthermore, SPR tasks appear to be well suited to addressing the aforementioned research questions, for the reasons discussed below.

Firstly, with regard to RQ 1 (To what extent can L2 case markings be learned under incidental learning conditions?), learning is operationalised with SPR tasks as differences in reaction times to grammatical versus ungrammatical utterances. Thus, SPR tasks are an indirect test of learning in that they measure the extent to which knowledge is deployed automatically. As noted previously, indirect tests have been argued to be more accurate measures of implicit knowledge than are traditional outcome measures, such as a GJT (see Section 3.2 above). This is because automaticity is considered to be a defining characteristic of implicit knowledge (see Section 2.3.3). However, it is important to acknowledge that, within alternative frameworks, specifically skill-acquisition theory, automaticity is indicative of proceduralised knowledge. Although the characteristics of procedural knowledge overlap with those of implicit knowledge, these constructs are not isomorphic (DeKeyser, 2009; see also Section 1.1.6). Thus, depending on one's theoretical orientation, different interpretations of the results from the SPR task in the present study are possible.

Another advantage of SPR tasks is that these tasks can be seen as more accurate in gauging the knowledge acquired as a result of incidental exposure (Hulstijn, 2003, 2013). As noted in Section 3.2, the concept of transfer-appropriate processing holds that we are more likely to remember what we have learned if there is a match between the conditions under which the material is learned and the conditions under which it is tested. In this regard, the SPR task is particularly advantageous in that the training and testing phases of the experiment are identical. Thus, in theory, the results of the SPR task might be a more accurate reflection of the learning that has taken place.

Concerning RQ 2 (What type of knowledge is developed as a result of incidental exposure; implicit or explicit?), another clear benefit of the SPR task is that there is no clear demarcation between the training and testing phases of the experiment. This is advantageous for a number of reasons. Firstly, previous research (Rebuschat, 2008) has demonstrated that participants can develop awareness as a result of rule-search behaviour when they know they are being tested. Thus, any implicit knowledge, as found in an SPR task, is less likely to be contaminated by explicit knowledge in this way. In addition, the use of an SPR task does not preclude the possibility of using a direct test at a later stage of the experiment. Thus, following in the tradition of the complementary use of direct and indirect tests (R. Ellis, 2005; Reingold & Merikle, 1988), the SPR task can be contrasted with the results of a direct measure of learning, thus facilitating a more refined look at the learning that has taken place during the experiment.

7.1. Methods

7.1.1. Participants

Forty-five native speakers of English at a university in the United Kingdom volunteered to take part in this study. Nineteen of these were excluded from the present study on the basis of having had previous experience of a language with a rich morphological system, such as Latin, Bulgarian or Russian. The remaining 26 participants (16 female, 10 male) were randomly assigned to one of two experimental groups: Group A (n = 13, 4 male, 9 female) or Group B (n = 13, 5 male, 8 female)[30]. None of the

[30] The issue of statistical power is addressed in the discussion section of this chapter.

participants selected for this study majored in linguistics or a foreign language. The ages of the participants ranged from 20 to 36 years (M = 24.74, SD = 4.86). The data from three participants were excluded due to irregular behaviour during the experiment (this point is elaborated upon in Section 7.2 below).

7.1.2. Research Ethics

Experiment 3 was approved by the research ethics committee of the Institute of Education, University of London. For each participant, informed consent was sought prior to beginning the experiment. Each participant was allowed to read through the information sheet describing the study, and was given an opportunity to ask questions or voice any worries or concerns about the experiment. Furthermore, as with the previous studies, it was stressed that participation in this study was voluntary, and that participants could withdraw at any point.

It should again be noted that the information sheet and consent form for this study were designed to mask the purpose of Experiment 3. Experiment 3 was presented as a study on sentence comprehension in a foreign language. Participants were told that this study was investigating the effects that 1) scrambled word order and 2) replacing an English word with a "nonsense word" had on their ability to understand the meaning of the sentence. Again, this deception was necessary to ensure the construct validity of the incidental learning conditions. It was only at the end of the debriefing session that participants were fully informed of the focus of this experiment, and they were given an additional opportunity to ask questions, voice concerns or withdraw their consent.

7.1.3. Stimulus Material

The stimulus domain of Experiment 3 was a slightly modified form of the stimuli used in Experiment 1. Although both Experiments 1 and 2 of the present thesis achieved some success in promoting learning of the target morphological forms, the stimuli from Experiment 1 were less complex than were the stimuli used in Experiment 2. Given the fact that Experiment 3 utilises a novel research methodology, and the difficulties Pilot Studies 1 and 2 had in finding a learning effect, it was decided that Experiment 3 would begin with a less complex system in the first instance. Once a significant learning effect is found, certain variables, such as the complexity of the stimuli, can then be modified to examine the effects that these modifications have on learning. A summary of the changes between Experiment 1 and Experiment 3 are presented in Table 7.1 below.

Table 7.1.
Summary of Changes to the Training and Testing Phases between Experiment 1 and Experiment 3

	Training Phase						Testing Phase
	Exemplars	Repetition	Total Exposure	Case markers	Syntactic Patterns	Training Task(s)	Testing Instruments
Experiment 1	48	3x	144 sentences (auditory only)	Nominative Accusative	2	1. Picture-matching 2. Elicited imitations (repeat sentence) 3. Elicited imitations (repeat foreign word in isolation)	1. 48-item auditory GJT (48 transfer items) 2. Confidence ratings 3. Source attributions 4. Debriefing questionnaire
Experiment 3	48	3x	144 sentences (visual only)	Nominative Accusative	2	1. Self-paced reading 2. Elicited imitations (repeat sentence)	1. Self-paced reading (control vs violation blocks) 2. 32-item multiple-choice task (32 transfer items) 3. Confidence ratings 4. Source attributions 5. Debriefing questionnaire

Training Set

The training set for Experiment 3 was identical to that of Experiment 1. It included 24 bi-syllabic, foreign nouns. Each of these nouns was used in two unique sentences in the training phase, one in the nominative case and one in the accusative case. In total, 48 unique sentences were used in the training phase of the experiment. All 48 sentences were repeated three times, once in each of the three training blocks, resulting in a total exposure of 144 sentences. Examples of these sentences can be seen in Table 7.2 below.

Table 7.2.
Descriptions and Examples of the Two Morphological Categories in Experiment 3

Morphological Category	Syntactic category	Example
Nominative	Subject	The britva cut David's face at the sink last night.
Accusative	Direct object	Peter used a britvu in the bathroom today.

In addition, as in Experiment 1, all stimulus sentences followed one of two syntactic patterns. One of these patterns followed a prototypical Subject-Verb-Object (S-V-O) word order, and the second followed an Object-Verb-Subject (O-V-S) word order. Further examples of the two syntactic patterns can be seen in Table 7.3 below.

Table 7.3.
Examples of the Two Syntactic Patterns Used in Experiment 3

Pattern	*Template*
O - V - S	$[[AP]_{TEMP} > [NP]_{OBJ} > [VP] > [NP]_{SUBJ} > [PP]]$
Nominative	*Last summer the grass ate the koza in the field.*
Accusative	*Last month the kasu opened Patrick with the key.*
S - V - O	$[[AP]_{TEMP} > [NP]_{SUBJ} > [VP] > [NP]_{OBJ} > [PP]]$
Nominative	*Last year the prodejna shipped goods to the shoppers.*
Accusative	*All week the builder took his vrtacku to work.*

As in the previous experiments, care was taken in the construction of the stimuli to ensure a balance between the nominative and accusative items, and between the two syntactic patterns, as well as to provide a counterbalance between these two considerations. In other words, the number of nominative items that followed an S-V-O pattern was equal to the number of nominative items that followed an O-V-S pattern, (please refer to Section 5.1.3 above for a fuller description of the construction of the stimuli). It is also important to point out that such counterbalancing amongst the items also serves to control against locality effects in sentence processing. Simply put, locality effects refers to the fact that the distance between related sentence elements impacts on the ability to establish their relationship during online processing (Bartek, Lewis, Vasishth, & Smith, 2011; Gibson, 1998, 2000). In the case of the present experiment, such locality effects might manifest in relation to the distance for which participants would be required to hold on to the grammatical information encoded within the inflectional marker in relation to stimulus sentences coded as *near* versus *far*. As can be seen in Table 7.4 below, sentences are labelled as either *near* or *far* with regard to distance. Within sentences coded as near, the foreign word falls in Segment 4, in a position immediately after the verb, which is a location typically reserved for the direct object of an English sentence. By contrast, the foreign word occurs in Segment 2 of sentences coded as far, a location typical of the subject of an English sentence. To illustrate how these sentences might be processed, let us examine Example 1 from Table 7.4 below. As participants read the first two segments of this sentence, *some time ago* / *the brana*, they might suspect that brana is the subject due to its position in the sentence. However, it is only after reading further, to Segment 4, *some time ago / the brana /*

blocked / the road that their understanding of the function of the foreign word, based on their incremental processing of the sentence, is either confirmed or disproved by the lexical clues contained in Segment 4. Thus, if participants are using the grammatical information encoded within the case ending, they would need to hold on to this information from Segment 2 until it could be verified in Segment 4. By contrast, in Example 2, *last week / the man / changed / the plenu / in the bathroom*, the foreign word appears in Segment 4. When the participant reaches the foreign word in the sentence, all of the information needed to interpret the function of the inflectional marker has already been encountered. This means that participants do not need to hold on to the information further; rather, their understanding of the sentence can immediately be corroborated or refuted. To summarise, sentences coded as *far* require participants to hold on to the grammatical information encoded within the inflectional marker until it can be verified at a later point in the sentence. Sentences coded as near do not require participants to do this; instead, the veracity of the participants' interpretation of the sentence based on the inflectional marker can immediately be identified.

Table 7.4.
Characteristics and Examples of Sentences Used in the SPR Task in Experiment

				Example sentences				
	Syntax	Case	Distance	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
1	S-V-O	Nominative	Far	Some time ago	the brana	blocked	the road	to town.
2	S-V-O	Accusative	Near	Last week	the man	changed	the plenu	in the bathroom.
3	O-V-S	Nominative	Near	This evening	the room	cleaned	the sluzka	with water.
4	O-V-S	Accusative	Far	This morning	the stuhu	tied	the man	to the gift.

As can be seen in Table 7.5 below, the SPR task in Experiment 3 was divided into four experimental blocks. The sentences used in Blocks 1–3 were identical, albeit randomised, for all participants. There were, however, two separate versions of Block 4: Version A and Version B. Half of the participants completed Version A, and the other half completed Version B.

Table 7.5.
Description of the Training and Testing Blocks of the SPR task in Experiment 3

	Training phase			Testing phase
	Block 1	Block 2	Block 3	Block 4 (Version A or B)
Stimulus Sentences	48	48 (identical to Block 1)	48 (identical to Blocks 1 & 2)	48 novel sentences
Comprehension Questions (CQ)	48 (CQ set 1)	48 (CQ set 2)	48 (CQ set 3)	48 novel CQs

In contrast to previous experiments in this thesis, which asked participants to match the meaning of the foreign word to one of two pictures displayed on a monitor, Experiment 3 included comprehension questions as part of the task in the training phase of the experiment. The decision to include these questions was made for a number of reasons. Firstly, with regard to the construct validity of the incidental learning conditions, the participants were told that this experiment was investigating issues related to understanding the meaning of sentences presented in an L2. Logically, then, a comprehension question seems to match the cover story given to participants. Secondly, comprehension questions assist in ensuring that the reaction time data gleaned from the SPR task are reflective of underlying cognitive processes. Analysis of the accuracy of CQs can help to ensure that participants are not simply clicking through the sentences, but are instead processing them for meaning; therefore, the reaction times registered for each segment of the sentence are more likely to represent underlying cognitive processes (Keating & Jegerski, 2015; Roberts, 2012b).

For this experiment, three comprehension questions were constructed for each of the 48 stimulus sentences used in the training phase of the experiment. As each stimulus sentence was repeated three times across this phase, this entailed participants having to answer a different comprehension question each time they encountered a particular stimulus sentence. In other words, although each stimulus sentence was repeated across the three training blocks, the comprehension questions were not. This was done on the assumption that changing the comprehension questions would help to ensure that participants maintained focus throughout the experiment.

The overarching principle in the construction of the comprehension questions was that these questions should support the cover story and be focused on the meaning of the stimulus sentence. In addition, a number of other considerations were taken into account in their development. Firstly, none of the comprehension questions included the foreign word contained in the stimulus sentence. In addition, for each stimulus sentence, the focus of the comprehension questions was balanced across three different components of the sentence. For example, if we look at the syntactic templates in Table 7.4 above, we can see that each sentence contains an adverbial time phrase (designated as [AP] in the template), as well as a prepositional phrase (designated as [PP] in the template). As such, for each sentence, one comprehension question concerned the adverbial phrase (when the action took place), a second question related to the prepositional phrase (where the action took place), and the final question concerned the action that took place in the stimulus sentence. The comprehension questions were balanced in this way to ensure that the comprehension questions were not responsible for directing participants' attention to one particular aspect of the stimuli, but were helping to ensure that participants attended to all parts of the stimulus sentences instead. Examples of the comprehension questions can be seen in Table 7.6 below.

Table 7.6.
Sample of Comprehension Questions (CQ) Used in the Training Phase of Experiment 3

Original Stimulus Sentence: *Some time ago David dug a studnu at his home*

	Focus	Example
CQ 1	Action	*Did David fix something?*
CQ 2	Time	*Did this happen recently?*
CQ 3	Location	*Did David do something at work?*

Furthermore, the answers to the comprehension questions were balanced across the experiment so that each foreign noun included equal numbers of comprehension questions with "yes" and "no" answers. For example, the three comprehension questions for the foreign noun *studnu* in the accusative case would elicit the answers yes, no and no, respectively. By contrast, the comprehension questions for studna, the nominative form, had the correct answers no, yes and yes, respectively. By balancing the comprehension questions in this matter, the total number of yes and no answers were equal with regard to case (nominative versus accusative), syntactic pattern (S-V-O versus O-V-S), for each foreign noun, within each training block and across the entire training phase of the experiment. A complete list of the stimulus sentences and corresponding comprehension questions can be found in Appendix G.

The fourth and final block (Block 4)[31] of the SPR task represented the testing phase of this stage of the experiment. Block 4 included 48 novel sentences that were not used in the first three blocks of the experiment. Block 4 included both grammatical and ungrammatical sentences, which were distributed across four stages within this block: a transition phase (*n*=16), a control phase (*n*=8), a violation phase (*n*=16) and a second control phase (*n*=8). This information can be seen in Table 7.7 below.

Table 7.7.
Distribution of Stimulus Sentences and Comprehension Questions across the Testing Phase of the SPR Task.

	Testing Phase			
	Block 4			
	Transition	Control 1	Violation	Control 2
Stimulus sentences	16 gram	8 gram	16 ungram	8 gram
Comprehension questions (CQ)	16	8	16	8

The transition and control phases contained only grammatical items, while the violation phase contained only ungrammatical items. As in the GJT used in previous experiments in this thesis, the grammatical and ungrammatical sentences were designed with the same considerations as the stimulus sentences used in the first three blocks of the SPR task. The grammatical sentences included the correct case marker. The ungrammatical items in the violation phase of Block 4 were generated by replacing the

[31] The procedure for the testing block of the SPR task was identical to that of the training blocks. This is discussed in more detail in the sections that follow.

correct case marking with the other case marker. Of the 16 ungrammatical sentences, eight were nominative and eight were accusative. The eight ungrammatical nominative items were created by replacing the correct nominative marker (-*a*) with the accusative marker (-*u*). The same procedure was followed to create the ungrammatical accusative items.

In addition, as noted above, there were two separate versions of Block 4: Version A and Version B. The difference between these is that the violation and control items were rotated between these two versions. In other words, the grammatical items used in the control blocks from Version A were made ungrammatical by changing the case marker. These items were then used in the violation block in Version B. The ungrammatical items from Version A were made grammatical by replacing the incorrect case marker with the correct one. These items were then used in the control blocks in Version B. Counterbalancing the test items in this way helps to ensure the internal validity of this study; specifically, differences in reaction times for grammatical and ungrammatical sentences are due to the case marker, and are not the result of another aspect of the sentence.

As noted previously, Experiment 3 also included a second measure of learning, in this case a multiple-choice task (M/C task). For this instrument, 32 novel sentences were constructed following the same considerations as those included in the SPR task. The sentences used in this task are comparable to the sentences used throughout the SPR task, as well as to the items included in the GJT in previous experiments in this thesis. Of the 32 items included in the M/C task, 16 were nominative and 16 accusative. An example of an item from the M/C test can be seen in Figure 7A below.

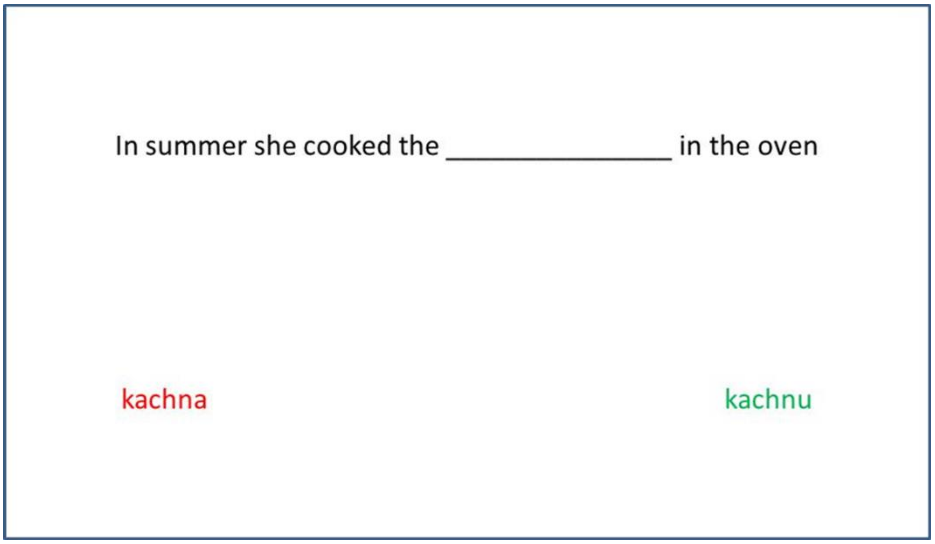


Figure 7A. Example item from the M/C Task of Experiment 3

7.2. Procedure

Experiment 3 took place in a private room on campus. The entire experiment was delivered via a Dell ThinkPad T410 laptop computer, using Superlab 4.5 (Cedrus Corp, San Pedro, CA) for the SPR and M/C tasks on a Cedrus model RB-834 response pad. In addition to providing greater ease of experimentation, response pads provide far more accurate reaction-time measurements than can be expected when reaction times are measured using a computer keyboard. Following the M/C task, participants completed a short debriefing questionnaire, followed by an oral interview. A depiction of the stages of Experiment 3 can be seen in Figure 7B below.

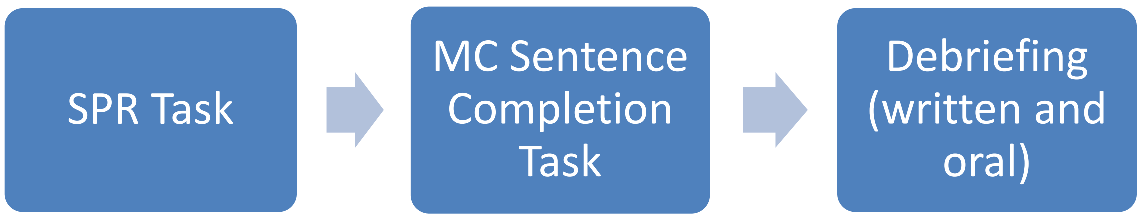


Figure 7B. Stages of Experiment 3.

SPR Task

Following common practice, this SPR task followed a non-cumulative linear design with a segment-by-segment presentation (Jegerski, 2014; see also Section 2.6 above). An illustration of how the SPR was presented can be seen in Figure 7C below.

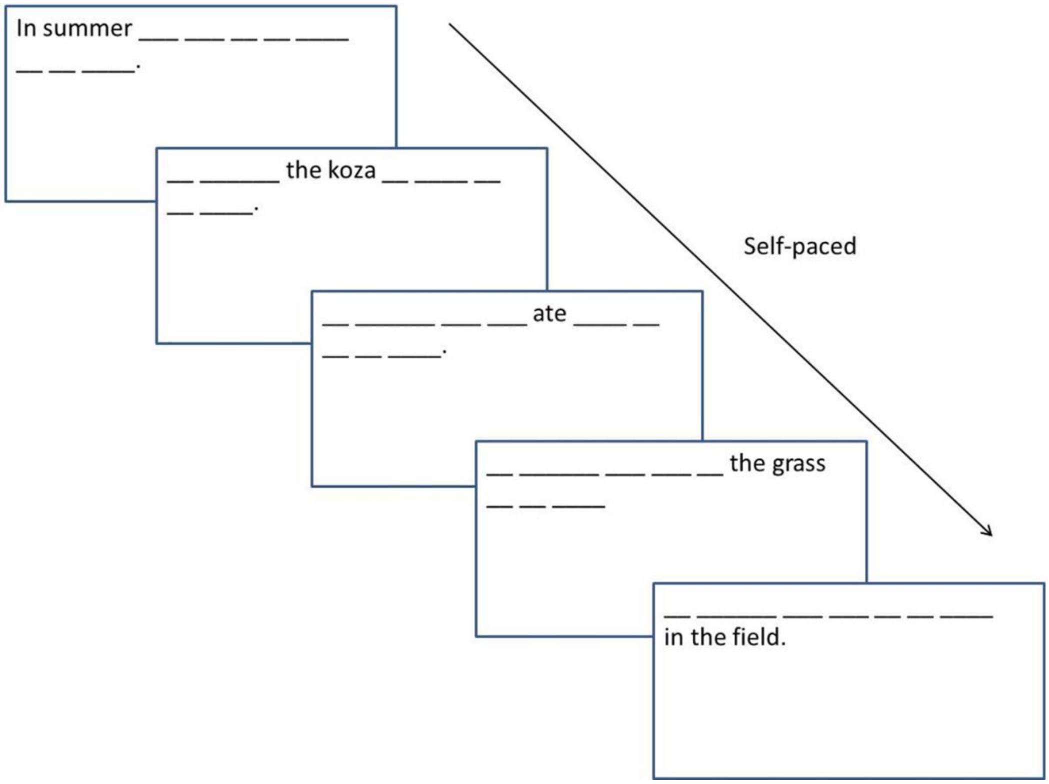


Figure 7C. Example of Slides Used in the SPR task in Experiment 3.

The procedure was identical for both the training and testing blocks of the SPR task. The instructions given to participants stated that they should read through the sentence quickly, but to take care that they understood it as they proceeded; in other words, participants were instructed not to just click through a sentence without understanding it. Their task was to read through each sentence (such as *Last summer / the grass / ate the koza / in the field*), and then to repeat it silently, from memory, when prompted by "*repeat*" on the monitor. In the initial piloting of the procedure, participants were asked to repeat the sentences aloud, as in the procedure in Experiments 1 and 2. However, the pronunciation of the foreign word was problematic as the participants had not encountered it auditorily, and several participants paused to rehearse the pronunciation of the foreign word when they encountered it in the SPR task. This threatened the internal validity of the SPR task, as the reaction times would be contaminated by issues connected with the pronunciation of the foreign word, rather than reflecting the cognitive processes connected with the comprehension of the sentence. Hence, the decision was made to have the participants repeat the sentence silently, rather than aloud. There were some drawbacks to this decision, primarily in that there was no way for the experimenter to ensure that participants were in fact repeating the sentences when prompted. To help to ensure that the participants were following the instructions provided, the experimenter observed whether the participants paused when prompted to repeat a sentence. Three of the participants (3/26) did not pause when prompted to repeat it (reaction time $\leq$ 500 ms) and continued to click through the prompt rapidly, even when reminded of the experimental procedure. Due to this deviation from the experimental procedure, the data from these three participants were excluded from further analyses.

After repeating the sentence silently, the participants pressed a key on the response pad and progressed to a yes/no comprehension question based on the sentence they had just read; for example, "Was the grass eaten by something?" Participants answered this question by pressing one of two clearly labelled buttons on the response pad. As feedback has been identified as a potential confounding variable, particularly due to its potential for raising awareness (Leow & Hama, 2013; Hama & Leow, 2010), no feedback was provided to participants regarding the accuracy of their answers. After answering the comprehension question, the next sentence began and the procedure was repeated for the next sentence in the sequence. Figure 7D illustrates this procedure, and Figure 7E below shows sample slides used in the SPR task.

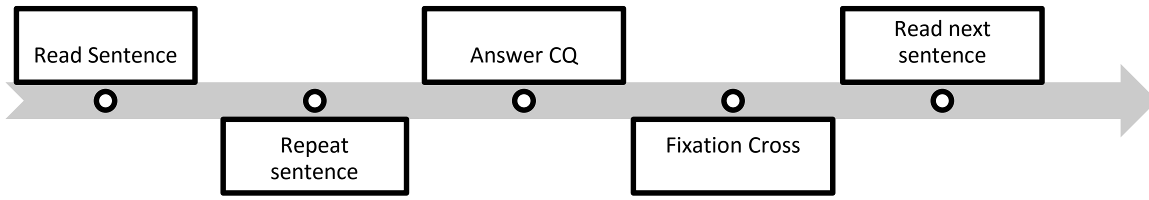


Figure 7D. Training procedure in Experiment 3.

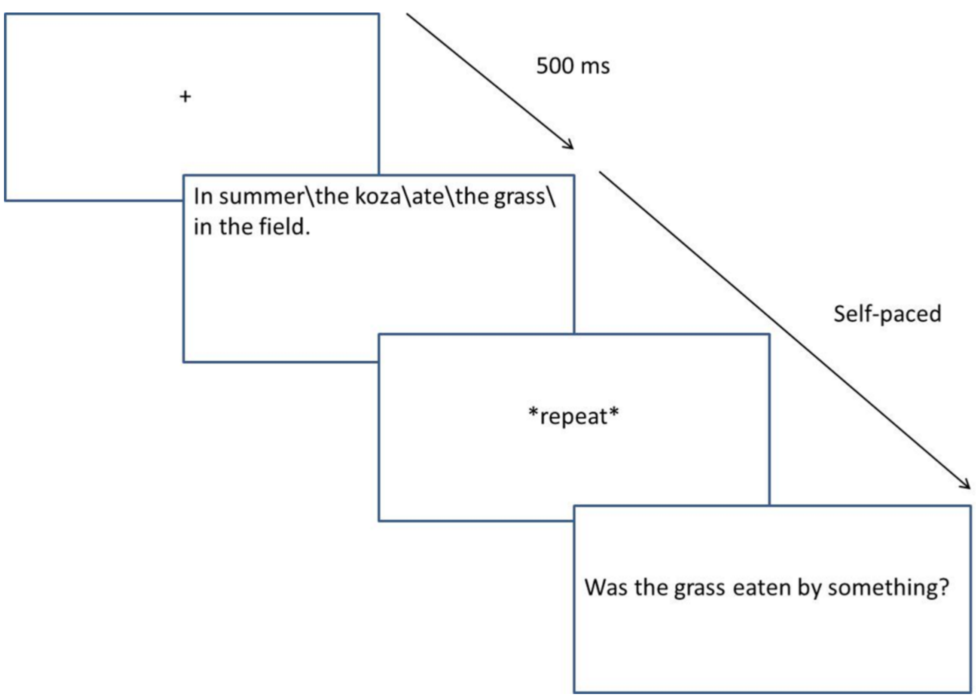


Figure 7E. Example of Slides in the SPR Task in Experiment 3.

Prior to the start of Training Block 1, participants read through the instructions for the task and completed four practice sentences in order to become familiar with the experimental procedure. Participants were also given the opportunity at this point to ask questions if they were unsure about any aspect of the experiment or the task they had to perform. Following the practice sentences, and the opportunity to ask questions, participants began Training Block 1 of the experiment. After the completion of each training block, participants were given a one-minute break. This was done to combat fatigue, on the assumption that such breaks would help to divide the experiment into more manageable chunks. Following the completion of Training Block 3, participants were given a final one-minute break before beginning the fourth and final block of the SPR task. The entire training phase, exclusive of the time spent on instructions, practice sentences and breaks, took about 40 minutes to complete.

As noted above, following Training Block 3, participants were given a one-minute break before beginning the Testing Block (Block 4) of the experiment. It is important to note that there was no demarcation between the training and testing phases of the experiments and, as noted above, that the procedure for the testing block was identical to that of the training blocks. To the participants, the Testing Block appeared to be simply a new set of sentences and comprehension questions.

Multiple-choice Task

Following the end of Block 4 in the SPR task, participants in the experimental group were informed that the sentences in the previous section were not arbitrary, but were part of a complex system. For each test sentence, participants were encouraged to decide, as quickly as possible, which of the two choices best completed the sentence. No feedback was provided concerning the accuracy of the participants' decisions. Alongside each test item, participants completed both confidence ratings and source attributions following each test item. See Figure 7F below for examples of slides that were used in Superlab for the testing phase of the experiment.

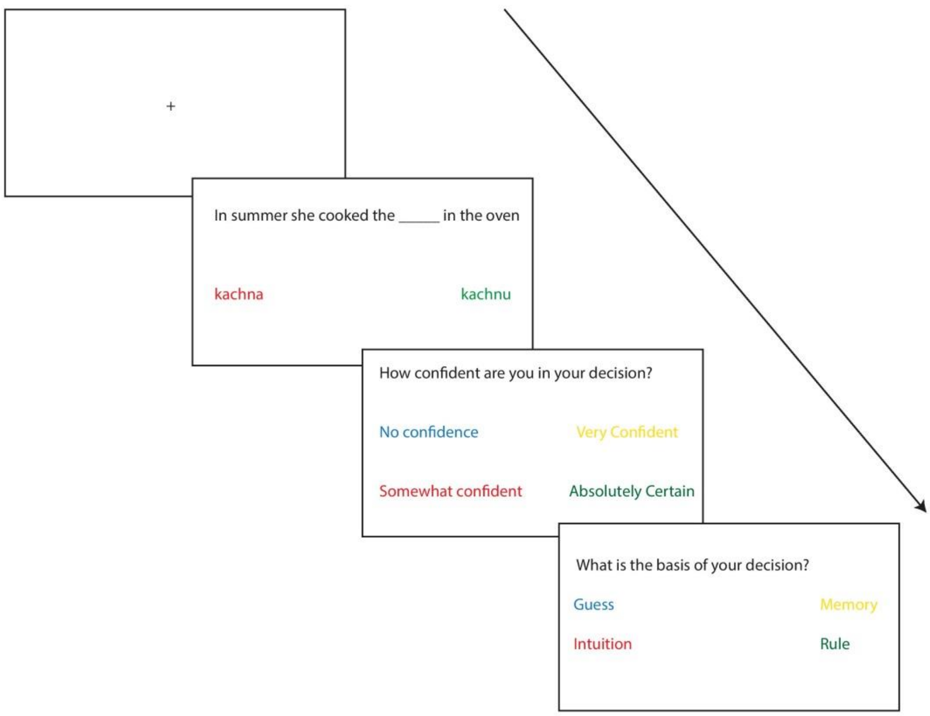


Figure 7F. Example of the Slides and Procedure for the M/C Task in Experiment 3.

Verbal Reports

Following the M/C task, participants were also prompted to describe, in both a written questionnaire and as part of a follow-up oral interview, any rules or patterns they might have noticed. In addition, the questionnaire for Experiment 3 included a question that asked if participants had been trying to engage in rule-search behaviour during the experiment. No such question was included in the previous experiments in this thesis, although this has been done in other studies investigating incidental learning (Robinson, 1997). A copy of the questionnaire for Experiment 3 can be found in Appendix I.

In the oral interviews, the researcher asked the candidates to elaborate on their responses to the written questionnaire. If the participants reported noticing the morphological markers, the experimenter asked the participants at what point in the experiment they remembered first becoming aware of them. If the participants were able to verbalise the underlying rule system, the experimenter asked them to elaborate at which point in the experiment they first thought of this system, as well as what made them become aware of the rules. Finally, if the participants remained unaware - in other words, if they were unable to state a morphological rule system - the researcher explained the underlying rule system and again asked the participants if they had figured this out or had any intuition about it at any point during the experiment.

7.3. Statistical Analyses and Data Treatment

As in Experiments 1 and 2, Experiment 3 utilises mixed-effects models to analyse the performance of the participants on both the SPR and the M/C tasks. All models were set up using the maximal structure advocated by Barr et al. (2013). As noted previously, this maximal structure requires that models include corresponding random effects for all fixed effects included within mixed-effects models. Following Cunnings and Sturt (2014), if the maximal model failed to converge, the random effect that accounted for the least variance in the data was removed until the model did converge. As was the case with the outcome measure in the previous experiments (the GJT), the M/C task in Experiment 3 generated categorical data in the form of correct / incorrect responses. Therefore, the M/C task was analysed using a logit mixed-effects model (Jaeger, 2008;

Linck & Cunnings, 2015). Statistical significance for the overall model fit was determined using the χ^2 statistic of likelihood ratio tests. Following the example provided by Baayen (2008), effect sizes for the logit-mixed effects model were calculated using the C index of concordance and the Hmisc package (Harrell & Dupont, 2015). C index values range from .5 to 1, with .5 representing no fit and 1 representing a perfect fit of the data. As noted in the preceding chapters, a C-index of .7 is interpreted as a moderate fit, .8 and above as a good fit, and .9 and above as an excellent fit for the data (Gries, 2013). Following Gelman and Hill (2007) and Linck and Cunnings (2015), significance of the fixed effects within the best-fitting model was interpreted as an absolute t-value ≥ 2.0. Further post-hoc Tukey tests were carried out to examine significant interaction effects with this model. Effect sizes for these tests were calculated using Cohen's *d*. Following Plonsky and Oswald's (2014) recommendations, Cohen's *d* of .40 is considered small, .70 medium, and 1.00 and higher as a large effect size.

In contrast to the M/C task, the SPR task elicited continuous data in the form of reaction times. To analyse these data, a linear mixed-effects model was constructed to examine the differences in reaction times for grammatical and ungrammatical items (Baayen et al., 2008; Linck & Cunnings, 2015). As with the logit mixed-effects models, statistical significance in building the linear model was determined using the χ^2 statistic of likelihood ratio tests, and significance for the fixed effects within the best fitting model was interpreted as an absolute t-value ≥ 2.0. Effect sizes in the form of R^2 were calculated using the MuMIn package in R (Barton, 2013). R^2 values, which range from 0 to 1, are an effect size estimate that quantifies the proportion of the variance in the data explained by the linear-effects model (Baayen, 2008; Cohen, 1988). The interpretation of R^2 values is intuitive, in that an R^2 value of 1 indicates that the model accounts for 100% of the data, whereas a value of 0 indicates that the model accounts for 0% of the data. As Experiment 3 is the first experiment in this series to elicit reaction-time data, the following section will provide a brief overview of some of the methodological issues related to the handling of reaction-time data prior to statistical analysis.

7.3.1. Treatment of Data in Reaction-time Research

Most statistical procedures, including mixed-effects models, assume that data are normally distributed (Osborne, 2005). This is of particular importance when working with reaction-time (RT) data, as they tend to be skewed and to contain outlying data points (Ratcliff, 1993; Whelan, 2008). Outliers, broadly defined, refer to reaction-time values that are generated by processes other than the one(s) under investigation (Ratcliff, 1993). For instance, in the case of the present study, slips of the hand, or momentary lapses in concentration, might result in unusually low or high reaction times that do not reflect the real reading times accurately, or the underlying cognitive processes under investigation (Jegerski, 2014; Jiang, 2012; Keating & Jegerski, 2015). As should be apparent, identifying and dealing with outlying values has been considered to be an important part of reaction-time research (see Jiang, 2012, for an overview).

Traditionally, outliers are identified as individual data points with extreme absolute values, either very high or very low, or data points that are extreme in relation to other data points in a particular data set. Treatment of outliers involves either removing outlying values from the data set or treating these values so that they do not skew the overall data set (see Leung & Williams, 2011, for an example in SLA research). It is important to pause here to differentiate between a priori data trimming and trimming as a result of model criticism. A priori data trimming, as the name implies, refers to treating data prior to committing the data to statistical analysis. For example, the mean and standard deviation may be calculated for a particular data set. A cut-off point is set, typically at 2 or 2.5 standard deviations from the mean. Values greater or smaller than this cut-off point are then discarded, or are treated by replacing the extreme value with the cut-off value (see Lachaud & Renaud, 2011; Ratcliff, 1993; Whelan, 2008 for overviews of a priori data treatment in psycholinguistic research).

One challenge in identifying outlying data points when working with reaction times is directly related to the fact that raw (untreated) RTs tend to be skewed towards shorter response times (Whelan, 2008). This is because there is a natural boundary at the low end of the distribution; in other words, there are physical limits to how fast it is possible to respond to a stimulus, but there is no such boundary at the high end. This results in a distribution with a positive skew, as exemplified by the raw RT data taken from the present SPR task. These data are illustrated in Figure 7G below.

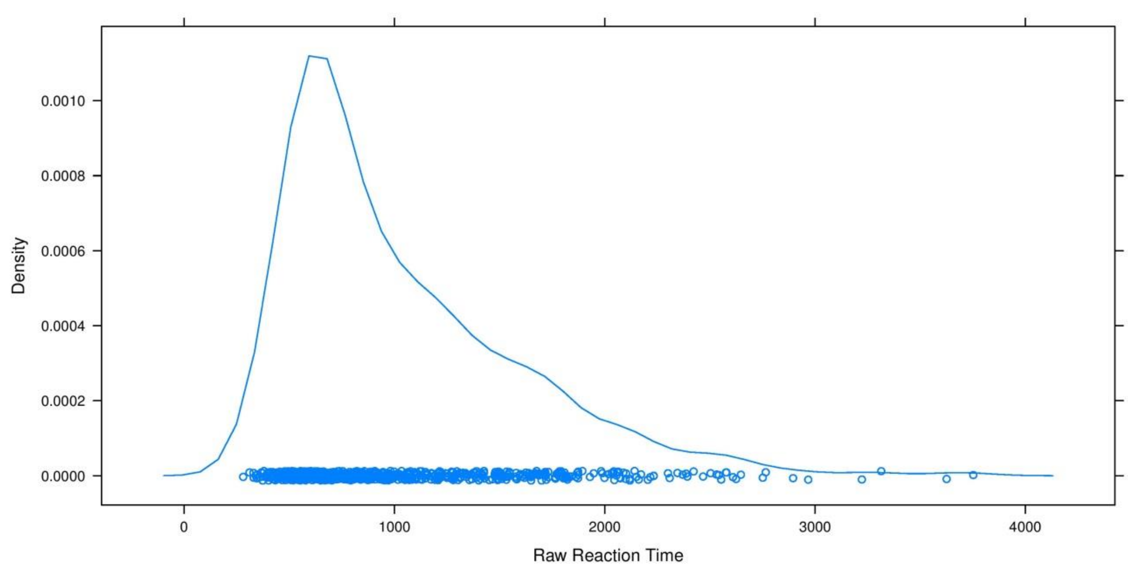


Figure 7G. Distribution of Raw Reaction Times from Experiment 3.

The problem with such a distribution is that long outliers "hide in the tail" (Ratcliff, 1993, p. 511). In other words, long outliers can have a dramatic effect on the central tendency of the data set to the point at which outliers become difficult to identify using the a priori treatment methods outlined above (Baayen, 2008; Baayen & Milin, 2010; Lachaud & Renaud, 2011). To address this issue, psycholinguists typically transform their data in order to arrive at a data set with a normal distribution, using either a logarithm transformation or a reciprocal one (see Box & Cox, 1964; Kliegl, Masson, & Richter, 2010; Osborne, 2002; Tabachnick & Fidell, 1996, 2001 for overviews)[32]. Tabachnick and Fidell (1996) argued that the shape of the distribution of the data set should determine the type of transformation used. For example, for positively-skewed distributions of shape as presented in Figure 7G above, they argued that a logarithmic transformation is preferred. Following transformation of the data, the researcher should be able to analyse the data set in order to identify and remove outliers from it.

Although the identification and treatment of outliers is considered vital for parametric statistics such as ANOVAs, mixed-effects models are generally robust in the face of outlying data points (Baayen, 2008; Baayen et al., 2008; Jegerski, 2014; Jiang, 2012; Keating & Jegerski, 2015), to the extent that some have argued that data trimming is unnecessary with these models (Jegerski, 2014; Keating & Jegerski, 2015). However, whether or not mixed-effects models are indeed impervious to the effects of outliers is a matter of debate (Baayen, 2008; Barr et al., 2013). What does appear to be important within fixed-effects models is checking that the assumptions of the mixed-effects model are met, specifically the assumption that the residuals of the fitted model are normally

[32] See also Gries (2013) and Pallant (2010) for instructions on transforming data in R and SPSS, respectively.

distributed (Baayen et al., 2008). Common procedures of data transformation and the screening of outliers help to ensure that this assumption is later met. It has been argued that, if the residuals of a model follow a normal distribution, outliers in the data set do not necessarily need to be removed. By contrast, if the residuals are non-normally distributed, it is only then that outliers should be identified and treated. This approach is referred to as model criticism, and it has been demonstrated to result in better model fit and fewer lost data points, than does traditional a priori data screening (Baayen & Milin, 2010). Therefore, a model criticism approach is adopted here.

7.4. Results

7.4.1. Verbal Reports

An analysis of the retrospective verbal reports revealed that eight of the 23 participants were able to verbalise the morphological rule system in some form, and seven of the eight were able to state the rule when prompted by the written questionnaire. The remaining participant answered "Yes, I thought of that" when the experimenter explained the rule. These participants are henceforth classified as "aware". Comments related to the subject and/or object of the sentence, as well as layman's explanations, such as the "when it ends in *-a* it does the action", were taken as evidence that participants were fully aware of the rule system. When asked when they had worked out the underlying rules, two of the eight said that it was while performing the SPR task, and the remaining six of the eight said it was during the final M/C task. Both participants who became fully aware during the SPR task were asked if they had noticed anything strange about any of the sentences towards the end of that task. When they answered no, it was explained that some of the sentences in the SPR task were ungrammatical as they carried the wrong morphological marker. Both of these participants claimed that they had not noticed any incorrect markers during the SPR task. One of these participants further commented that, "Maybe that's why some of the sentences seemed more difficult to understand." Finally, when asked if they had engaged in rule-search behaviour during the experiment, all eight of the aware participants answered in the affirmative, although four of them stated that they only began looking for rules during the M/C task, not during the SPR task.

The remaining 15 participants (of 23) remained unaware of the underlying rule system, and are henceforth classified as being "unaware". When prompted, all 15 of these participants were able to identify both of the inflectional markers included in the present study (*-a* and *-u*). This was taken as evidence that they had, at a minimum, noticed the surface features of the input at some point during the experiment. Three participants offered incorrect rules related to word order. One of these three participants stated that the *-a* marker was used at the beginning of a sentence, and the *-u* marker at the end, regardless of word order.[33] The other two participants stated that one of the markers was used with normal word order (S-V-O), and the other marker was used with reverse word order (O-V-S). When asked to guess the rule behind these markers, eight of the 15 unaware participants guessed gender. One participant guessed that the *-a* marker was used with singular nouns, and *-u* with plurals. Another participant stated that these markers were used with nouns, but was unable to elaborate further. The remaining four unaware participants claimed they had no idea, even when prompted to guess. One of these participants remembered noticing the case endings during the SPR task, but "didn't

[33] The experimenter followed this up to ensure that this participant was not referring to the subject/ object distinction.

think they were important". When asked if they had engaged in rule-search behaviour during the experiment, only two of the 15 unaware participants claimed to have done so in the M/C task, but none of these participants indicated that they were looking for rules during the SPR task. Concerning rule-search behaviour, one participant did not look for rules because of being "too busy concentrating on meaning" during the SPR task. In summary, as in previous experiments in this thesis, all of the participants in this study demonstrated awareness at the level of noticing (Schmidt, 1990). However, in contrast to previous experiments, a number of participants developed awareness of a higher order, which could be classified as awareness at the level of understanding following Schmidt's (1990) framework.

7.4.2. SPR Task

7.4.2.1. Comprehension Questions

Firstly, in order to gauge the degree of difficulty that participants had in comprehending the stimulus sentences in the SPR task, descriptive statistics were calculated to examine the extent to which participants were able to answer the comprehension questions (CQs) correctly throughout the experiment. Such an analysis would not only provide information about the degree of difficulty of the stimulus sentences, but would also provide an indication as to whether the participants were simply clicking through the sentences without trying to comprehend them. The results of this analysis revealed that the participants answered 76.75% (*SD* = 17.64%) of the comprehension questions correctly across the entire SPR task. The performance of the aware participants (*M* = 77.17%, *SD* = 12.09%) was then compared to the performance of the unaware participants (*M* = 79.93%, *SD* = 4.55%) via a logit mixed-effects model. The null model included accuracy as a binary outcome, with random intercepts for subjects and items. A second model was constructed, which incorporated group (aware versus unaware) as a fixed effect, and as a random slope by item. A likelihood ratio test revealed that the inclusion of group did not lead to a significantly improved fit - χ^2 (1) = 1.34, p =0.63, C = .52 - indicating no significant difference between the performance of the aware and unaware participants in the comprehension questions during the SPR task. Accuracy on the comprehension questions was further broken down across the three training blocks of the study, as well as for control and violation items in the testing phase of the experiment. These results are presented in Table 7.8 below.

Table 7.8.
Percentage Accuracy of Participants on Comprehension Questions across the Training and Testing Blocks of Experiment 3

Group		Training Phase			Testing Phase	
		Block 1	Block 2	Block 3	Control	Violation
Total	*M*	77.14	82.16	77.17	80.74	82.17
	SD	10.28	9.92	8.34	11.42	13.51
Aware	*M*	75.52	80.21	73.70	80.47	80.04
	SD	14.98	13.38	9.14	14.80	19.38
Unaware	*M*	78.00	83.20	79.03	80.89	83.31
	SD	6.33	7.13	7.14	9.12	8.72

To determine whether the variation across the training and testing blocks was statistically significant for either the aware or unaware groups at any point in the experiment, an additional logit mixed-effects model was constructed, which included a random effect for time (Block 1, Block 2, Block 3, Control Phase and Violation Phase). Although this inclusion was significant - χ^2 (3) = 11.09, p =0.04, C = .64 - this effect did not interact with group (aware versus unaware). This indicates that the performance of the aware and unaware participants was not significantly different across any of the training blocks. The significant main effect for time is perhaps best explained by the slightly reduced accuracy in Block 1 and Block 3 for aware and unaware participants alike during the training phase. The slightly reduced rates of accuracy in Block 1 could best be explained by the fact that the participants were still becoming used to the stimuli. Due to the repetitive nature of the experiment, the reduced rate in Block 3 could be attributed to either to fatigue or boredom.

7.4.2.2. Reaction-time Data

In order to examine the response times, it is important to identify in which segment of the sentence we might expect to find differences in the reaction times to grammatical and ungrammatical items. As can be seen in Table 7.9 below, the foreign word always appears in either Segment 2 or Segment 4 of the sentence. Furthermore, the foreign word always functioned as either the subject or as the object of the sentence. Regardless of whether the foreign word appeared in Segment 2 or Segment 4, the reader would not be able to identify whether the inflectional marker was correct until reaching Segment 4.

Table 7.9.
Sample Sentences Illustrating Pre-critical, Critical and Post-critical Regions of Stimulus Sentences

	Regions				
	Pre-critical	Pre-critical	Pre-critical	Critical	Post-critical
	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
1	Some time ago	the brana	blocked	the road	to town.
2	Last week	the man	changed	the plenu	in the bathroom.
3	This evening	the room	cleaned	the sluzka	with water.
4	This morning	the stuhu	tied	the man	to the gift.

Thus, if the participants were processing the sentence incrementally, and were using the grammatical information coded within the inflectional markers in real time, then Segment 4 would be the first segment in which differences in reaction times to grammatical and ungrammatical utterances would be registered. Thus, Segment 4 can be referred to as the *critical region*, in that it is the point in the sentence at which the grammaticality of the utterances can be realised. However, in addition to examining the difference in reaction times within the critical region, it is also important to examine the data that both precede and follow the critical region. These regions are referred to as the *pre-critical region* and the *post-critical region*, respectively (Jiang, 2012). The reaction-time data from the pre-critical region serve as a baseline, or control, to which data from the critical region can be compared. It is to be expected that no difference in reaction times to grammatical and ungrammatical sentences will be observed in the pre-critical region (Jiang, 2012; Jegerski, 2014). For example, if the same differences in reaction times between grammatical and ungrammatical utterances are found in both the critical region and the pre-critical region, then it is likely that these differences are an artefact of the testing set, or are a result of the speed at which the entire sentence was read. However, if there is no such difference in the pre-critical region, then it can be said with more certainty that the demonstrated differences in reaction times in the critical period are due to the grammaticality of the utterance.

In addition to analysing data from the pre-critical region, it is also important to analyse the post-critical region, or the region that follows the critical region. This analysis is important due to the possibility of spillover effects in reaction times (Traxler & Tooley, 2008). Spillover effects refer to the phenomenon when processing is delayed or is incomplete within a particular region and, as a result, this processing is carried over to the next word or segment in the sentence (see Keating & Jegerski, 2015 for a brief discussion). In other words, the reaction times in a particular region are influenced "not only by the problems in the current display but also by any backlog or processing that may have built up in the buffer" (Mitchell, 1984, p. 76).

Previous research that has utilised the SPR technique (Traxler & Tooley, 2008) has found evidence of spillover in post-critical regions of the sentence. Thus, an analysis of the post-critical position, Segment 5, might reveal whether participants' demonstrated some form of delayed processing of grammatical and ungrammatical items. It is also noteworthy that it is important to control for spillover on an item-by-item basis. For example, the speed at which a participant processes the pre-critical region in a particular sentence might influence the speed at which he or she processes the critical region. Thus, more recent studies within psycholinguistics have attempted to control for this potentially confounding variable by incorporating spillover effects as a fixed effect in their mixed-effects models (Hofmeister, 2011; Hofmeister, Jaeger, Arnon, Sag, & Snider, 2013.

7.4.2.3. Transformation of Reaction-time Data

Prior to any statistical analysis, the reaction-time (RT) data were initially screened for unrealistic values. Following Baayen (2008; see also Baayen & Milin, 2010), reaction times of less than 400 ms and greater than 4,000 ms were removed from the data set. This resulted in the removal of 12 data points from nine participants (eight values less than 400 ms, four values greater than 4,000 ms), which represented approximately 1.7% of the total data points. Descriptive statistics were calculated to examine the average RTs of aware and unaware participants in the critical period across the training and testing phases of the experiment. As can be seen in Table 7.10 below, the reaction times of the aware participants were consistently greater than were those of the unaware participants.

Table 7.10.
Average Response Time (ms) for Aware and Unaware Participants across Training and Testing Blocks for Segment 4

		Training Phase			Testing Phase		
Group		Block 1	Block 2	Block 3	Transition	Control	Violation
Aware	m	1262.55	1161.90	1242.94	1275.48	1237.61	1323.97
	sd	621.28	591.22	714.99	559.98	545.91	627.96
Unaware	m	967.75	981.94	947.92	921.45	846.62	892.06
	sd	575.42	592.90	594.53	562.97	456.65	560.18

To examine whether this difference extended to grammatical and ungrammatical items during the testing phase of the experiment, additional descriptive statistics were calculated for the control and violation blocks, as well as for the various sub-components, which will be analysed later in the linear mixed-effects model. These results can be seen in Table 7.11 below.

Table 7.11.
Reaction Times for All (Aware and Unaware) Participants for Control and Violation Items in Segment 4

All Participants		Total	Nom	Acc	SVO	OVS	Close	Long
Control	M	985.15	980.65	989.65	904.36	1065.94	1057.35	912.85
	SE	28.10	39.50	40.07	33.20	44.64	42.17	36.48
	SD	538.97	535.76	543.59	450.32	605.53	572.02	494.82
Violation	M	1026.07	1052.79	999.35	982.59	1069.55	1172.81	879.33
	SE	30.904	44.55	42.87	43.03	44.26	49.34	34.53
	SD	592.83	604.35	581.51	583.63	600.34	669.25	468.33
Aware Participants		Total	Nom	Acc	SVO	OVS	Close	Long
Control	M	1237.61	1200.53	1271.45	1122.23	1349.75	1319.83	1182.16
	SE	48.28	62.73	73.63	60.95	72.60	73.068	62.97
	SD	545.91	501.83	589.06	487.49	580.81	584.54	503.78
Violation	M	1323.97	1302.19	1321.69	1254.81	1369.06	1434.28	1139.59
	SE	52.16	73.016	71.67	72.86	71.13	80.44	55.30
	SD	627.96	584.13	573.39	582.85	569.07	643.52	442.40
Unaware Participants		Total	Nom	Acc	SVO	OVS	Close	Long
Control	M	846.62	863.38	839.36	788.17	914.57	933.51	769.23
	SE	31.34976	47.18	41.36	34.98	51.55	48.11	38.98
	SD	456.65	517.90	453.016	383.19	564.66	527.03	427.04
Violation	M	892.06	919.78	827.34	837.30	909.82	1071.36	755.86
	SE	35.11	52.42	46.57	48.55	50.73	56.59	38.95
	SD	560.18	574.23	510.17	531.84	555.71	619.93	426.65

Following common practice in the psycholinguistic literature, the plot of the distribution of reaction times was generated to examine the distribution of the untransformed RTs. As this plot demonstrated a positive skew, the data were log-transformed, and were re-examined for normality. Plots of both the overall log-transformed reaction times, as well as the log-transformed reaction times by individual participants, revealed that the data were now normally distributed. These plots can be found in Appendix K.

7.4.2.4. Analysis of Critical-region Data

Performance on the SPR task was examined via a linear mixed-effects model. To begin, two analyses were carried out to control for potentially confounding variables. The reader will recall that the testing phase began with a transition block, followed by a control block of eight items, a violation block of 16 items and a second control block of eight items. In order to determine if reaction times differed between the two control blocks, a linear mixed-effects regression model was built using the lme4 package (Bates et al., 2014) in R version 3.1.2 (R Development Core Team, 2008). Within this model, the log-transformed reaction times for control items only were modelled as a continuous dependent variable together with cross random intercepts for subject and items. A second model, which incorporated time (control block 1 versus control block 2) as a fixed effect and as a random slope by participant, was then built. The addition of time as a fixed effect did not result in a significantly improved model when comparing control items that came before and after the violation block - $\chi2$ (1) =0.43, p =.51, R^2=.67 - indicating that there were no significant differences in reaction times between items that occurred in the two control blocks. Thus, the items from the two control blocks were merged into a single group for subsequent models. Next, and more crucially, analyses were carried out to examine whether the two versions of the test, Versions A and B, which were counterbalanced among the participants, impacted on reaction times. To this end, another linear model was constructed in which test version was added as a fixed effect to the previous model, with the addition of a test version as a random slope by item. This version was then compared to a null model that included only the crossed random intercepts of subject and items. The results of a goodness of fit analysis indicated that the addition of a test version did not lead to a better fit for the data: χ^2 (4) = 42.18, p = 0.70, R^2=.69. This null result indicates that participants performed similarly on the two versions of the test, and thus these versions do not account for the variance in the data.

At this point, further models were built to determine the fixed effects that had a significant relationship with the dependent variable, namely differences in reaction times to grammatical and ungrammatical items. To do this, it was necessary to include the fixed effect of grammaticality in the null model. If grammaticality were excluded from the null model, then maximum likelihood analyses would indicate relationships between fixed effects and overall reaction times, and not with that of the dependent variable. First, a null model was constructed, which analysed the log-transformed reading times from the entire testing set (both control and violation items) in terms of the fixed effect of grammaticality (grammatical versus ungrammatical items). This model also included crossed random intercepts for subjects and items, as well as a random slope for grammaticality by participant (Barr et al., 2013). As noted previously, such a random slope by participant would allow the model to account for variance in grammatical versus ungrammatical items on a participant-by-participant basis, and a random slope by item would account for such variance on an item-by-item basis. Additional models were built for each fixed effect. Each of these models included a variable, such as case, as a fixed effect as well as a random slope (Barr et al., 2013). These models were then compared individually against the null model using the $\chi 2$ statistic. If the model that included the fixed effect was significant against the null model, this result was interpreted as indicating that the fixed effect in question had a significant relationship with the dependent variable, and should be included in subsequent analyses. If the result was nonsignificant, this was interpreted as no significant relationship, and that this fixed effect could be excluded from the models that were to follow. These data are presented in Table 7.12 below.

Table 7.12.
Summary of Maximum Likelihood Ratio Tests on Predictor Variables in SPR Task

Predictor Variable	χ^2	df	p	R^2
Case	20.51	9	.02*	.70
Syntax	12.14	9	.21	.69
Locality	77.14	9	< .001***	.71
Awareness	9.76	4	.05*	.69
M/C Task	5.70	4	.22	.69

As can be seen in Table 7.12 above, the results of these comparisons indicated that the fixed effects of case, locality, and awareness showed a significant relationship with the dependent variable. Therefore, a new model that incorporated these three fixed effects was constructed. In addition, following the maximal model as suggested by Barr et al. (2013), case, distance and awareness were included as random effects: case and distance were included as random slopes by participant, and awareness as a random slope by item. A maximum likelihood ratio test indicated that this model was significant against the null model, χ^2 (23) = 126.85, p < .001, and accounted for approximately 75% of the variance in the data, R^2=.75. Further comparisons indicated that the inclusion of all of these effects was warranted in the model, and the removal of any resulted in a decreased R^2 value. Multicollinearity was formally assessed using Cohen's kappa. This resulted in a value of 1.76 which, according to Baayen's (2008) guidelines, indicates a very low level of collinearity between the fixed effects in the model. Thus, this model is henceforth referred to as the best-fit model, the results of which are presented in Table 7.13 below.[34]

[34] The residuals of this model were inspected and found to be approximately normally distributed. A Q-Q plot of this distribution can be found in Appendix K.

Before examining the data in more detail to determine which fixed effects reached significance within the best-fit model, it is important to eliminate the possibility that spillover from the pre-critical region is the source of the variance in the critical region[35]. Spillover values were calculated on an item-by-item basis, following the R model provided by Jaeger (https://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/; see also Hofmeister, 2011). To determine whether spillover from the previous region could account for the variance in the previous model, a new null model, which analysed the residuals from the previous model with crossed random intercepts for subjects and items, was created. Spillover was then entered as a fixed effect into a new model with a random slope by subject as a random effect. The results indicate that the inclusion of spillover as a fixed effect did not lead to a significantly better fit for the data: $\chi 2$ (1), =0.66, p =0.42, R^2=.39. This result indicates that spillover from the pre-critical region is not a significant explanatory variable for the variance in RTs in the critical region.

At this point, it is worth examining a summary of the best-fit model, as presented in Table 7.13 below. Summary statistics allow for an examination of the extent to which individual fixed effects, as well as interactions between fixed effects, contribute to the model. Following Gelman and Hill (2007) and Linck and Cunnings (2015), significance here is interpreted as an absolute *t*-value ≥ 2.0.

[35] The post-critical region was also examined for any spillover effects from the critical region. No such effects were found. Please refer to Appendix K3 for more information regarding analyses carried out on the post-critical region.

Table 7.13.
Results of the Best-fitting Linear Mixed-effects Model for the Critical Region of the SPR task.

	Fixed Effects			Random Effects	
				By Subject	By Item
Parameter	Estimate	*SE*	*t*	*SD*	*SD*
Intercept	6.87	0.07	92.31*	0.33	0.11
Gram	0.04	0.03	1.34	0.03	
Case	-0.02	0.04	-0.40	0.15	–
Locality	-0.23	0.05	-4.80*	0.17	–
Awareness	0.47	0.15	3.25*	–	0.06
Gram: Case	-0.04	0.05	-0.85		
Gram: Locality	-0.12	0.05	-2.30*		
Gram: Aware	<.-001	0.04	-0.01		
Gram: Case:Aware	-0.01	0.08	-0.14		
Gram: Locality:Aware	-0.01	0.04	-0.26		

Note. All factors were coded using contrast coding, as follows: Locality (−.5 = short, .5 = far), Case (−.5 = Nominative, .5 = Accusative), Grammaticality (−.5 = control item, .5 = violation item) formula: RT~gram*locality*case*awareness + (locality*case*gram|subject) + (aware|item).*|t| > 2.0, indicating a significant effect (Gelman & Hill, 2007).[36]

As can be seen in Table 7.13 above, several of the main effects reached significance in accounting for the variance in the reaction-time data. Firstly, the data indicate that aware participants responded more slowly than did unaware participants. In addition, participants responded more slowly to "near" sentences (instances in which the foreign word occurred after the verb, such as *this morning the man changed the plenu in the bathroom*), than they did to "far" sentences (instances in which the foreign word appeared before the verb, such as *some time ago the brana blocked the road to town*).

[36] Contrast coding has been argued to aid in model convergence as well as reduce the impact of multicollinearity in mixed-effects models (Linck & Cunnings, 2015).

The reader will probably notice at this point that the fixed effect of grammaticality did not achieve significance as a main effect within this summary. This result, however, does not suggest that the grammaticality of the utterance did not impact on the results. It is important to re-emphasise that "a main effect need not be significant if it is involved in interactions which are significant" (Baayen, 2012, p. 166). In other words, although the differences in RTs between grammatical and ungrammatical items might not have been sufficiently robust to achieve significance as a main effect, the interaction between the grammaticality of the sentence and another main effect might account significantly for the variance in the data. As can be seen in Table 7.13 above, there was a significant interaction between grammaticality and locality. This result suggests that differences in reaction times to grammatical and ungrammatical utterances can be found in sentences in which the foreign word appears in the critical position (near) versus when an English word appears in the critical position (far). An interaction plot was generated to examine this relationship more closely. This plot can be seen in Figure 7H below.

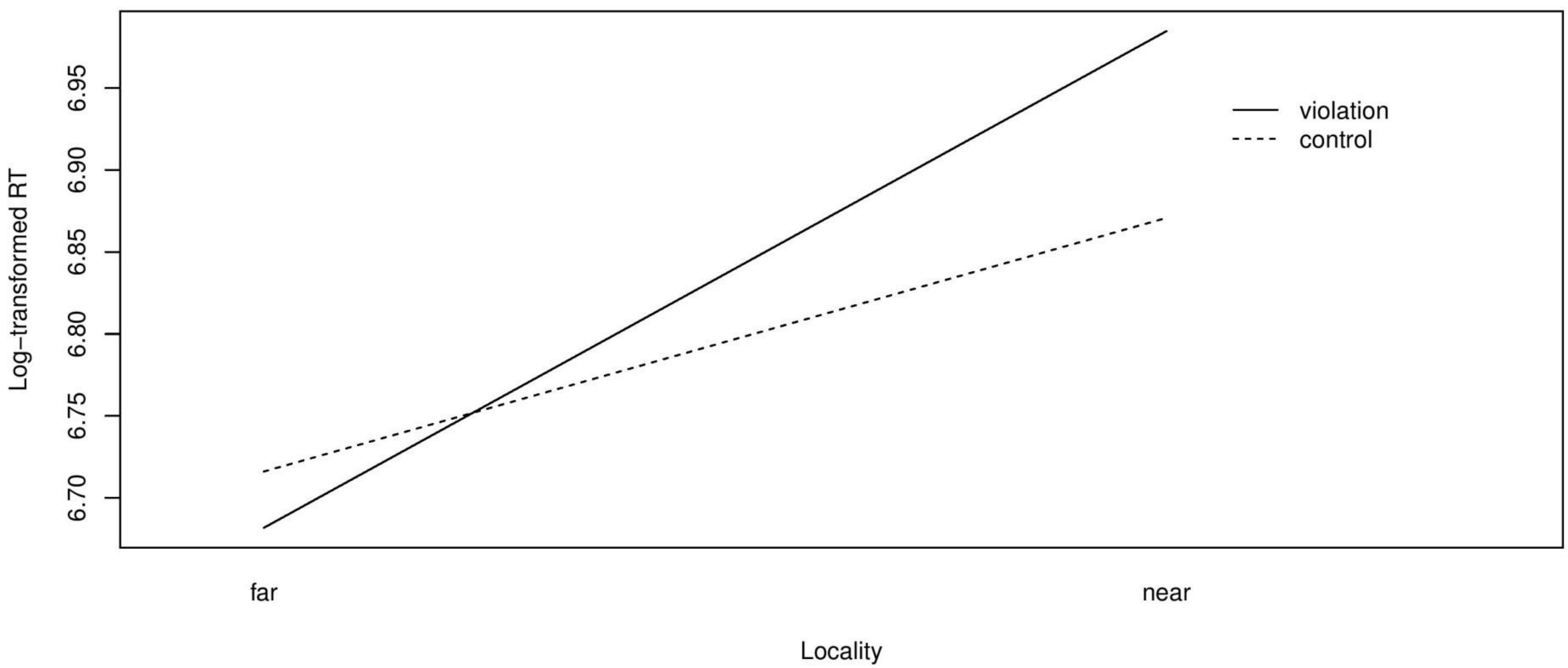


Figure 7H. Plot of Interaction between Locality, Grammaticality and Reaction Times.

As Figure 7H above clearly illustrates, reaction times to grammatical and ungrammatical utterances were nearly identical for far sentences. However, for near sentences, reaction times to violation items were considerably greater than were reaction times to control items. This result suggests that participants' sensitivity to grammatical violations within the foreign word was mediated by the position of the foreign word within the sentence. To examine this result further, post-hoc Tukey tests were calculated using the GLHT command of the R package "multcomp" (Hothorn et al., 2008). These tests confirmed that the differences in reaction times to near violation and control items were statistically significant, B = -. 09, SE=.04, z = -2.57, p=.04, d=.41. However, differences in reaction times to grammatical and ungrammatical utterances for far sentences were nonsignificant, B = -.02, SE = .04, z = -.57, p =. 93, d = .07.

Finally, it is important to reiterate that, although awareness was significant as a main effect, it was not significant as an interaction with grammaticality. This indicates that the reaction times of aware participants were longer on average than were those of unaware participants, both for grammatical and for ungrammatical items. However, the fact that awareness did not interact with grammaticality indicates that aware and unaware participants behaved similarly with regard to differences in reaction times to grammatical and ungrammatical utterances. In other words, awareness does not appear to be an explanatory variable for the learning effect demonstrated in the SPR task.[37]

Further analyses of the pre-critical and post-critical regions found no significant effects for grammaticality. In other words, the data indicate that differences in reaction times to grammatical and ungrammatical utterances did not extend beyond the critical

[37] However, it should be noted again that this nonsignificant finding might be attributed to a lack of statistical power. This point is discussed in more detail below.

region of the sentence. Descriptive statistics and summaries of the mixed-effects models used to examine the pre-critical and post-critical regions can be found in Appendix K.

7.4.3. M/C Task

An analysis of the participants' performances in the M/C task revealed that they were able to classify 57.65% of the items correctly (*SD* = 17.17%). Descriptive statistics were also generated to compare the performance of aware versus unaware participants. Aware participants were able to identify the correctly inflected noun with 69.14% accuracy (*SD* = 20.21%), whereas unaware participants were only able to do so at a rate of 51.88% (*SD* = 12.15%). Further descriptive statistics were generated for the performances of aware and unaware participants across the various sub-components of the M/C task. These results are presented in Table 7.14 below.

Table 7.14.
Performance on Sub-components of the M/C Task.

			Case		Syntax		Locality	
All Participants (n=23)		Total	Nom	Acc	SVO	OVS	Near	Far
	M	57.65	54.19	61.31	55.67	56.71	59.48	56.01
	SD	17.17	20.96	22.60	16.99	19.17	16.99	18.96
	SE	3.58	4.37	4.71	3.54	4.00	3.54	3.95
Aware Participants (n=8)		Total	Nom	Acc	SVO	OVS	Short	Long
	M	69.14	63.94	74.34	62.41	66.94	71.78	66.50
	SD	20.22	25.61	22.89	20.30	23.33	20.58	23.15
	SE	7.15	9.05	8.09	7.18	8.25	7.18	8.18
Unaware Participants (n=15)		Total	Nom	Acc	SVO	OVS	Short	Long
	M	51.67	49.92	53.42	52.08	51.26	52.92	50.42
	SD	12.15	17.21	22.23	13.08	16.34	13.75	15.58
	SE	3.14	4.47	5.75	3.38	4.22	3.55	4.02

Analyses were carried out to determine which fixed effects had a significant relationship with the dependent variable, namely performance in the M/C task. To this end, a series of logit mixed-effects models were built. For these analyses, a null model that analysed performance in the M/C task as a binary outcome in terms of subject and item as crossed random intercepts was constructed. Additional models were then built for each fixed effect. Each of these models included the variable, such as case, as a fixed effect as well as a random slope (Barr et al., 2013). As was done with the previous analyses in this thesis, these models were then compared individually against the null model using the $\chi 2$ statistic. As noted previously, if the model that included the fixed effect was significant against the null model, this result was interpreted as indicating that the fixed effect in question had a significant relationship with the dependent variable, and should be included in subsequent analyses. If the result was nonsignificant, this was interpreted as no significant relationship, and that this fixed effect could be excluded from the models that were to follow. The results of these comparisons can be seen in Table 7.15 below.

Table 7.15.
Summary of Likelihood Ratio Tests on Predictor Variables in the M/C Task

Predictor Variable	χ^2	df	p	C
Case	21.28	3	<.001***	.71
Syntax	0.56	3	.91	.64
Distance	0.79	3	.85	.64
Aware	5.26	3	.02*	.66

As can be seen in the table directly above, the results of these comparisons indicated that the fixed effects of case and awareness showed a significant relationship with performance on the M/C task. Therefore, a new model that included both of these variables as fixed effects was constructed. In addition, following the maximal model as suggested by Barr et al. (2012), the fixed effect of case was included as a random slope by participant, and awareness as a random slope by item. A maximum likelihood ratio test indicated that this model was significant against the null model, χ^2 (7) = 30.41, p < .001, C=.73. Further comparisons indicated that the inclusion of both case and awareness as fixed effects was warranted in the model, and the removal of either fixed effect did not lead to an improved model fit. Multicollinearity was formally assessed using Cohen's kappa. This resulted in a value of 1.52 which, according to Baayen's (2008) guidelines, indicates a very low level of collinearity between the fixed effects in the model. Thus, this model of the data from the M/C task is henceforth referred to as the best-fit model, the results of which are presented in Table 7.16 below.

Table 7.16.
Results for the Best-fitting Logit Mixed-effects Model Examining Performance on the M/C task

	Fixed Effects				*Random Effects*	
Parameters	Estimate	*SE*	*z*	*p*	By subject *SD*	By items *SD*
Intercept	.55	.14	3.84	<.001***	.55	.20
Case	-.65	.35	-1.87	.06^{+}	1.00	–
Awareness	.93	.28	3.36	<.001***	–	.13
Aware: Case	-.37	.18	-2.06	.04*		

Note. Factors were coded using contrast coding, as follows: Case (−.5 = Nominative, .5 = Accusative), Awareness (-.5 unaware, .5=aware), formula: "correct~case*awareness + (case|subject) + (aware|item), family=binomial, control=glmerControl (optimiser="bobyqa")".

As can be seen in the summary in Table 7.16 above, there are a number of factors within this model that were significant in accounting for performance in the M/C task. Firstly, awareness (whether or not participants became aware of the morphological rule system) was significant at the *p* <.001 level, indicating that the differential performance of aware and unaware participants was significant. Although case only approached significance as a fixed effect (*p* =.06), there was a significant interaction between case and awareness in explaining accuracy in the M/C task. This suggests that either the aware or the unaware participants might have performed differentially on test items that targeted the nominative versus the accusative markers.

To examine this result further, post-hoc Tukey tests were carried out using the GLHT command of the R package "multcomp" (Hothorn et al., 2008). First, these analyses revealed that the aware group was significantly more accurate for accusative items than for nominative items, albeit with a small effect size; B = -.74, SE = .28, z = -2.62, p = .04, d =.38. By contrast, the unaware participants' performance on nominative versus accusative items did not differ significantly, B = -.15, SE = -.19, z =-.83, p = .83, d =.10. Statistical comparisons between the aware and unaware participants' performances indicated that the aware participants significantly outperformed the unaware participants on both nominative (B =-.58, SE = .27, z = -2.12, p = .14, d =.61) and accusative items (B = -1.16, SE=.29, z = -3.97, p = <.001, d =.88). Taken together, these results clearly indicate that the performance of the aware participants, particularly regarding the accusative test items, was the driving factor behind the overall learning effect.

7.4.3.1. Subjective Measures of Awareness

As the M/C task in Experiment 3 resulted in a significant learning effect, it is possible to analyse the results of the subjective measures of awareness. As discussed in previous chapters, subjective measures of awareness are designed to determine whether resulting knowledge is implicit or explicit in nature (see Section 3.5.3 for a discussion of these measures).

Source attributions. As can be seen in Table 7.17 below, an analysis of the source attributions revealed that participants relied largely on their intuition rather than on explicit rule knowledge in their judgements. As indicated by the results of the one-sample *t* tests, participants' performances were significant across three of the source attributions: intuition, memory and rule. However, their performances were not significant when attributing their decisions to *guess*. This means that the *guessing criterion* of implicit structural knowledge was not met, and thus the data from the source attributions point towards the knowledge reflected in the M/C task as being explicit in nature.

Table 7.17.
Accuracy (%) and Number of Responses for All Participants across Confidence Ratings and Source Attributions.

	Accuracy (%)	Number	*p*-value
Confidence Rating			
No confidence	55.06	154	.20
Somewhat confident	59.62	420	<.001
Very confident	53.57	139	.40
Absolutely certain	72.73	23	.03
	Accuracy (%)	Number	*p*-value
Source Attribution			
Guess	51.68	147	.68
Intuition	56.87	280	.02
Memory	58.38	221	.02
Rule	70.79	88	<.001

Confidence ratings. As shown in Table 7.17 above, participants' performances were significantly better than chance levels when indicating that they were somewhat confident or absolutely certain about their decisions. However, as noted in Section 3.5.3, the criterion for implicit knowledge as measured by confidence ratings is the zero-correlation criterion (Dienes & Scott, 2005), which holds that knowledge can be considered implicit if there is no relationship between reported levels of confidence and accuracy (for example, in a GJT). Following the common statistical procedure in the cognitive sciences (Rausch et al., in press; Sandberg et al., 2013; Wierzchon et al., 2012), confidence level was analysed as a predictor variable within a logit-mixed effects model. To do this, a null model that included accuracy as a fixed effect along with crossed random effects of subjects and items was constructed. A second model was then constructed, which was identical to the null model, but which also included confidence as a fixed effect, as well as a random slope for both subjects and items. These models were then compared using a likelihood ratio test. The results of this test revealed that the confidence level was significant in accounting for variance in the dataset (χ^2 (8) = 20.50, p = 0.01, C = .67), indicating a relationship between confidence level and accuracy in the M/C task. Thus, it can be inferred that the zero-correlation criterion was not met, and that the knowledge represented in the M/C task was explicit in nature.

In summary, the results of the M/C task indicated that learning had taken place. However, this learning was limited to participants who had become aware of the target case-marking system. Analyses of subjective measures of awareness provided further evidence that the knowledge-driving performance in the M/C task was explicit in nature.

7.5. Discussion

The results of Experiment 3 support the findings from Experiments 1 and 2, in that the data indicate that participants can develop some knowledge of L2 case markings after a minimal amount of exposure, without feedback, and as a result of incidental exposure. In contrast to Experiments 1 and 2, Experiment 3 employed two measures of learning: an indirect measure in the form of reaction times to grammatical and ungrammatical utterances via a self-paced reading task, and a direct measure in the form of a multiple-choice sentence-completion task. With regard to the relationship between learning and awareness in this experiment, an analysis of the retrospective verbal reports indicates that successful performance in the M/C task was driven by participants who were able to demonstrate verbalisable rule knowledge of the target morphological structures. This result was also corroborated by the subjective measure of awareness, which provided further evidence that the knowledge demonstrated in this task was explicit in nature. By contrast, the learning demonstrated in the SPR task did not appear to be linked to verbalisable rule knowledge, suggesting that participants, both aware and unaware, were utilising more automatic knowledge within this task.

Research Question 1: To what extent can L2 case markings be acquired as a result of incidental exposure?

7.5.1.1. SPR Task

Analyses of the data from the SPR task also found evidence that learning of the target morphological structures had taken place. These data were analysed using a linear mixed-effects model with crossed random slopes for subjects and items. The real object of interest from the SPR task is whether participants would demonstrate knowledge of the inflectional markers through differences in reaction times to sentences that contained a foreign word with the correct case ending (grammatical sentences) versus sentences that contained a foreign word with an incorrect case ending (ungrammatical sentences). Such evidence was found in the form of a significant interaction between the fixed effects of grammaticality and locality. In other words, participants demonstrated learning in the form of extended reaction times to ungrammatical case-endings when the foreign word fell within the critical region (near sentences.). No such differences in reaction times were found, however, when the foreign word was located in the pre-critical region (far sentences). An example of a near versus a far sentence can be seen in Table 7.3 above.

This finding is intriguing, and it could be accounted for in a number of different ways. Firstly, differences in performance on near versus far sentences might be attributed to the phenomenon known as locality effect. As discussed in Section 7.1.3 above, locality effects refer to the fact that a greater distance between linguistically-connected sentence elements increases the difficulty in processing the relationship between these elements in real time (Bartek et al., 2011). Far sentences require that the grammatical meaning encoded in the inflection marker be carried from Segment 2, a pre-critical region, to Segment 4, a critical region. By contrast, no such distance exists in near sentences. Thus, one possible explanation for the results here is that the participants had developed knowledge of the grammatical information encoded in the case endings, but were not able to draw upon this knowledge in far sentences due to difficulties in retaining this information over the long-distance dependency. This finding appears congruent with the implicit learning literature, which has demonstrated that participants are better at learning regularities of adjacent items than they are with relation to non-adjacent items (Cleeremans & McClelland, 1991; Creel, Newport, & Aslin, 2004; Gómez , 1997, 2002; Howard, Howard, Dennis, & Kelly, 2008; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010; Pacton & Perruchet, 2008; Remillard, 2008; van den Bos, Christiansen, & Misyak, 2012).

A different, but perhaps complementary, explanation for the differences in performance on near versus far items might concern the shallow processing hypothesis (Clahsen & Felser, 2006; see also Section 2.9.3.2 above) and/or the lexical preference principle of VanPatten's model of input processing (VanPatten, 2004b; see also Section 2.9.3.1 above). The shallow processing hypothesis (Clahsen & Felser, 2006; Clahsen et al., 2010) holds that L1 speakers process input deeply, using the entire range of lexical and grammatical clues available to them as they parse input. By contrast, L2 learners process input "shallowly", in that they often rely on lexical and other non-structural clues, rather than on grammatical clues, such as might be encoded within a case-marking system. Similarly, the lexical preference principle of VanPatten's input processing model states that learners typically utilise lexical means to comprehend a sentence, rather than grammatical information. If we make the assumption that the participants in the present study were utilising both lexical and grammatical means to comprehend the sentences, but that the learners were relying primarily on lexical means, the shallow processing hypothesis and lexical preference principle appear to provide plausible explanations for the results of the present study. To elaborate, participants' primary form of understanding the meaning of the sentence was through lexical means. The use of inflectional morphemes to establish subject-object relations within the sentence was secondary. Sensitivity to grammatical violations was only seen when the grammatical morpheme conflicted with the interpretation of the sentence that the participants had derived from lexical and non-structural clues. In near sentences, such a conflict was possible. However, in far sentences, the inflectional morpheme occurred in the pre-critical region, where the participants had not yet worked out the meaning of the sentence. In this sense, an ungrammatical morpheme could not serve to confirm or conflict with their interpretation of the meaning of the sentence. It is important to stress that, although appearing to account for the difference in performance between near and far items, such an account is speculative in the absence of direct data. Furthermore, it is not entirely clear whether the learning effect attributed to near items might be partly attributed to the position of the foreign noun, i.e., whether the foreign noun was positioned before or after the verb as well as whether the sentence followed an S-V-O or O-V-S word order. Future research might investigate this issue more directly statistically via a larger data set, or methodologically by adding think-aloud protocols or simulated recalls to the methodology employed in the present experiment. Such instruments might provide more direct evidence of the cognitive processes that underlie the differences in performance demonstrated here.

Another interesting finding from the SPR task was the lack of a relationship between awareness and the learning demonstrated on near items. In contrast to the M/C task, in which learning was confined to participants who had developed awareness at the level of understanding of the target inflectional system, the analysis of the SPR task did not indicate such a relationship. It should be reiterated that there was a fixed effect for awareness. However, this effect indicates that the reading times for aware participants were longer for both grammatical and ungrammatical items, although this effect did not interact with grammaticality. In other words, there was no significant difference in performance between aware and unaware participants in the SPR task. To frame this result in a slightly different manner, awareness did not appear to play a part in the learning effect demonstrated in the SPR task. This finding is important, and it is discussed further in the section below.

Finally, it is worth questioning whether the reaction times in the critical period in Experiment 3 are necessarily reflective of the decomposition of the foreign words.

Psycholinguistic research in morphological processing (Clahsen & Felser, 2006; Clahsen et al., 2010; Gor & Jackson, 2013; Marsden, Altman, & St. Claire, 2013) has employed priming techniques, such as lexical decision tasks, to investigate the extent to which learners process inflected forms. One of the challenges in this area of research is that the stem of a word can act as a prime, as can the inflection. This issue is relevant to the present study due to the fact that the SPR task measured the reading time for the foreign word in its entirety, not the inflection in isolation. To address this issue, psycholinguistic research investigating morphological priming has typically utilised nonce words in the form of made-up words that obey the phonotactic rules of the language under investigation (Hahne, Mueller, & Clahsen, 2006), or has included completely novel words in the testing phase of experiments (Marsden et al., 2013). Thus, any priming effects observed are argued to be the result of inflection, as the stem will not have been encountered previously. Although future research would undoubtedly benefit from more fine-grained data than the SPR task is capable of providing, there are a number of reasons to believe that the reaction times reflect processing of the case endings. Firstly, as noted in the methodology section above, all of the foreign words included in the testing phase were novel; thus, it was not possible for the reaction times to reflect recognition of a previously-encountered stem. Secondly, as noted in Section 7.7 above, the stems of the foreign words for the control items and the stems for the violation items were rotated among the participants. In lieu of a control group, the rotation of the control and violation stems among participants increases the confidence that the reaction times are reflective of the processing of morphological inflections, and are not merely artefacts of individual test items.[38]

In summary, the fact that participants demonstrated differences in reaction times to grammatical and ungrammatical case markers in near sentences provides some evidence that the participants were engaged in online processing of the inflectional markers, that they had acquired some knowledge of the function of the nominative and accusative markers, and that they were utilising this knowledge automatically as they parsed the sentences in real time.

7.5.1.1. M/C Task

As was the case with the SPR task, the results from the M/C task indicated that learning had taken place. Data from the M/C task were analysed by generating descriptive statistics, and by using a logit mixed-effects model with crossed random intercepts for subjects and items. The logit mixed-effects model returned significant main effects for awareness (aware versus unaware participants), case (nominative versus accusative case), and a significant interaction between awareness and case. To illustrate this, with regard to overall performance, participants were able to identify a foreign word with the correct case ending with 57.65% accuracy (*SD*=17.17%). This performance is roughly similar to the overall performance in the GJT task in Experiments 1 and 2 (~56% accuracy). However, in contrast to the previous experiments, the learning effect in Experiment 3 appeared to be driven by aware participants; in other words, participants who were able to verbalise the underlying rule system. While the aware participants' performance was at roughly 70 per cent for the M/C task (*M* = 69.14%, *SD* = 20.22), the

[38] It should be noted that variance due to individual test items and subjects is accounted for within mixed-effects modeling. Thus, possible item effects were further controlled for via statistical analysis.

performance of the unaware participants (participants who were unable to verbalise the underlying rule system) was scarcely indistinguishable from chance performance (M = 51.67, SD = 12.15). Furthermore, similar to the previous experiments, this main effect was driven by the aware group's performance regarding accusative sentences (M = 74.34, SD = 22.89), compared to the performance of unaware participants for accusative items (M = 53.42, SD = 22.23). It is also important to note that, unlike the results from previous experiments, learning in Experiment 3 did not appear to be limited to accusative items. Aware participants were able to supply the nominative marker correctly in 63.94% of cases (SD = 25.61). Again, this is in contrast to the performance of the unaware participants, whose performance does not appear to indicate any learning of the nominative marker (M = 49.92, SD = 17.21).

The lack of learning amongst unaware participants on the MC task in Experiment 3 is interesting, given that Experiments 1 and 2 demonstrated learning in a similar task (a GJT), and that none of the participants in Experiments 1 and 2 were able to verbalise the underlying rule structure. In other words, the participants in Experiments 1 and 2, who would be classified as unaware according to the criteria in Experiment 3, showed learning in a GJT, whereas the unaware participants in this experiment did not show learning in an M/C task. One possible explanation for the lack of learning demonstrated by unaware participants in the M/C task might be connected with the difference in modalities between the first two experiments and Experiment 3. As noted previously, Experiments 1 and 2 were presented completely auditorily, as none of the stimuli were presented visually, whereas Experiment 3 was presented completely visually, as none of the stimuli were presented auditorily. It seems plausible that the auditory GJT in Experiments 1 and

2 encouraged participants to draw upon more implicit knowledge in the form of intuitive judgements. The fact that Experiment 3 was presented visually might have promoted the use of more explicit knowledge to complete the task. As aware participants had developed higher levels of explicit knowledge in the form of metalinguistic knowledge, they were able to complete the task with a high degree of accuracy. Unaware participants, on the other hand, had not developed sufficient explicit knowledge, and were therefore unable to perform above chance levels of performance. This interpretation is supported by comparing the proportions of the source-attribution data of Experiments 1 and 2 with the results from Experiment 3. On average, the participants in Experiments 1 and 2 attributed approximately 30% of their decisions to memory or rule, the two choices indicating higher levels of awareness (Serafini, 2013). By contrast, participants in Experiment 3 attributed approximately 50% of their judgements to memory or rule, suggesting a much higher reliance on explicit knowledge than was the case in the previous experiments. Furthermore, as noted previously, the M/C task as used in Experiment 3 is a direct measure of learning which, it has been argued, draws more heavily on explicit knowledge. This is in contrast to the indirect measure (the SPR task), which is considered to be more reflective of implicit knowledge.

A finding of Experiment 3 that is in line with the results from Experiments 1 and 2 concerns performance regarding items targeting the nominative and accusative markers. In Experiments 1 and 2, it was hypothesised that greater amounts of learning of the accusative markers were connected with the salience of these markers in the input; specifically, that the accusative marker (-*u*) was more morphophonologically distinct, which resulted in more noticing (Schmidt, 1995) and deeper processing (Clahsen &

Felser, 2006) during the training phase of the experiment. If, however, morphophonological salience was the sole explanatory variable for the learning patterns demonstrated in Experiments 1 and 2, this effect should have been nullified in Experiment 3, in which the input of the training and testing phases was presented visually. This was not the case. One explanation for this continued pattern of learning is that salience remains an important factor. For example, if one accepts the argument that, orthnographically, English words that end in *-u* (as in adieu, bayou, menu and milieu), are more likely to appear to be of French origin, or "foreign" to native speakers of English than those that end with the vowel *-a* (for example, tea, area, arena and mascara), it seems plausible that the morphotypological characteristics of the foreign words might have led to an increase in the attentional resources being devoted to the accusative marker. As type frequency in the input has been implicated in the extent to which participants are able to generalise grammatical knowledge to novel items (see Bybee & Thompson, 2000 and N. Ellis & Collins, 2009 for discussions of the effects of type versus token frequency on L2 learning), such a skewed distribution of noticing the accusative marker rather than the nominative marker could have impacted on the learning of the nominative and accusative forms.

It is also important to point out that the performance in the M/C task could also be explained as reflecting bias in favour of the accusative marker. This bias is evidenced directly from the differential results for test items targeting the nominative and accusative markers in the M/C task. To elaborate, the M/C asked participants to read a sentence (for example, *in summer she cooked the ________ in the oven*), and to then choose between two options, a foreign word with a nominative marker, or the same foreign word with an accusative marker (such as *kachna*/*kachnu*). The test items were balanced so that the accusative ending was correct for half of the test items, and the nominative ending for the other half. Thus, if a participant indicated that the accusative ending was correct for all test items, then his or her results would show 100% accuracy for accusative items, but 0% accuracy for nominative items. Although the difference in performance between nominative (~64%) and accusative (~74%) items for aware participants[39] was slight, the fact that performance was higher for accusative items indicates that participants were more likely to choose the accusative ending than they were the nominative ending.

There are a number of reasons that participants might have shown bias towards the accusative marker. One explanation might concern the morphotypological characteristics of the case endings, as discussed above. If the accusative marker were more salient in the input in the training phase of the experiment, it could be argued that the participants noticed the accusative marker with more regularity than they did the nominative marker. When we consider that repeated exposure has been shown to enhance feelings towards stimuli (Zajonc, 1968), increased noticing of the accusative marker might have led to increased feelings of appropriateness, manifested in the form of intuitive judgements in the M/C task. Yet another possibility lies in the form of an innate bias based on the L1 of the participants. As discussed in Chapter 4, several researchers have argued that the accusative is the default in English (Emonds, 1995; Schütze, 2001). In this sense, the bias towards the accusative marker could be the result of an L1 transfer effect. Yet another possibility is a processing bias related to the L1 of the participants in

[39] ~54% nominative vs ~61% accusative for both aware and unaware participants.

this study. In this regard, the participants might have applied an S-V-O parsing strategy, thus interpreting the first noun in the utterance as the subject, regardless of the actual word order of the stimuli. Evidence for this parsing strategy might be seen in the difference in reaction times for SVO and OVS items in the critical region for both violation and control items in the SPR task (see Table 7.11). Slower reaction times to OVS items suggests that participants were devoting more cognitive resources to the processing of the critical region, perhaps due to a potential garden path effect (e.g., Roberts & Felser, 2011) in their parsing of the sentence. This interpretation finds further theoretical support in issues related to blocking and learned attention within the literature. Certain theoretical frameworks, such as the associative-cognitive CREED (N. Ellis, 2007b; N. Ellis & Wulff, 2015), have argued that, through implicit learning of the L1, the nature of the language-learning mechanism is optimised according to the characteristics of the L1. In this sense, attention to certain grammatical components is learned based on the characteristics of the L1 (N. Ellis, 2006a, 2007b, 2008; N. Ellis & Sagarra, 2010a, 2010b, 2011). The literature also documents that the negative impact of learned attention extends to the learning of morphosyntax, where redundancy of meaning can result in overshadowing and blocking of the morphological markers (N. Ellis, 2006a; Goldschneider & DeKeyser, 2001; Gor, 2010; Marsden, 2006; VanPatten, 2014). It follows that, if the L1 of the participants carries a default processing bias, it could be possible that this bias could carry over into their processing of L2 input.

Research Question 2: What type of knowledge is acquired: implicit or explicit?

An interesting point of departure for this section is the fact that approximately one third of the participants (eight of 23) were able to demonstrate verbalisable rule knowledge of the target morphological structures via the retrospective verbal reports at the end of Experiment 3. In this sense, these participants clearly developed explicit knowledge, given that verbalisability has long been the main criterion for explicit knowledge (Rebuschat, 2013). This is in contrast to the results from Pilot Studies 1 and 2, as well as to the results of Experiments 1 and 2, in which none of the participants were able to verbalise the underlying morphological rules at the end of the experiment. The simplest explanation for why participants in Experiment 3 developed awareness, and those in the previous experiments did not, lies with the modality of the experiments. As noted previously, all the stimuli in Experiment 3 were presented visually, whereas all the stimuli in the previous experiments were presented auditorily. Although little research has addressed the relationship between modality and the development of implicit and explicit knowledge directly (Rebuschat, 2013), what little research there is has indicated that written input is more amenable to controlled processing, which is typically associated with the use of explicit knowledge, whereas auditory input is more conducive to more automatic processing, typically associated with the use of implicit knowledge (Bialystok, 1982). It seems plausible, then, that the visual stimuli promoted explicit processes amongst the participants, such as rule-search behaviour. These explicit processes thus led to increased levels of awareness, as well as to a number of participants "cracking the code" and developing an understanding of the underlying rule system.

An alternative account might be that the aware and unaware participants exhibited different patterns of behaviour throughout the experiment. As noted above, aware participants demonstrated longer reaction times for both the pre-critical and critical regions across the testing phase of the experiment. Importantly, the analysis of reaction times across the training phase of the experiment showed a similar pattern of extended reading times for aware participants. One explanation for this could be connected to the shallow processing hypothesis (Clahsen & Felser, 2006), in that the differences in reaction times between aware and unaware participants represent differences in the depth of processing (Leow, 2015b). To be more specific, aware participants may have processed the segments of the sentence more deeply, which was reflected in longer reaction times. Another explanation might concern rule-search behaviour, particularly because the aware participants might have been more active in trying to identify patterns in the input. This resulted in longer reading times as they searched for rules. At their core, such interpretations are of the chicken or egg variety, especially regarding whether the longer reading times led to increased awareness as the experiment progressed, or whether increased awareness from the outset was reflected in longer reading times. Such a finding is intriguing, and warrants further investigation. Future research might utilise additional measures, such as eye-tracking or perhaps simulated recall, which would assist in coming to a fuller, more informed understanding of any differences in the online behaviour of aware versus unaware participants.

Shifting the focus to whether the exposure phase resulted in implicit knowledge, explicit knowledge or a combination thereof, measures of awareness associated with the M/C task and the SPR task provide two contrasting, but complementary, points of view. The results of the M/C task appear to be relatively straightforward in providing evidence for the development of explicit knowledge. This interpretation is gleaned from a variety of sources. First, the M/C task was a direct test of learning (see Section 3.5.4 above). The reader will recall that direct tests are tests that ask participants to make use of their knowledge, and that the results of direct tests have been argued to be more reflective of explicit rather than of implicit knowledge (R. Ellis et al., 2009; Godfroid et al., 2015). Second, analyses demonstrated that the overall learning effect demonstrated in the M/C task was driven by aware participants, namely participants who were able to verbalise the underlying morphological rule structure. In other words, the learning displayed in the M/C task was the result of participants who demonstrated high levels of awareness of the target structures. Finally, with regard to the subjective measures of awareness, neither of the two criteria of implicit knowledge, specifically the guessing criterion or the zero-correlation, was met. Thus, the subjective measures of awareness provide further evidence of explicit knowledge. These results appear to sit well with the findings from the previous experiments in this thesis, as well as with those of previous research systems (Brooks & Kempe, 2013; Grey et al., 2014; Robinson, 2002), which indicated clear links between awareness and the learning of L2 case-marking systems.

In contrast to the M/C task, a number of indicators suggest that performance in the SPR task was reflective of more automatic, implicit knowledge. With regard to the nature of the SPR task itself, it is an indirect test of learning in that it does not require participants to apply their knowledge. Instead, it measures learning indirectly via differences in reaction times to grammatical and ungrammatical utterances. In this sense, the SPR task can be argued to be more transfer appropriate (Hulstijn, 2013) to incidental learning research, in that it is a more sensitive measure of unconscious knowledge (Reingold & Merikle, 1988). Further support that the results of the SPR task reflect some implicit knowledge comes from the nonsignificant results for predictor variables entered into the linear mixed-effects model. In contrast to the results from the M/C task, the linear mixed-effects model did not identify significant differences between the performances of aware and unaware participants. In other words, there appeared to be a dissociation between metalinguistic awareness and learning, as demonstrated in the SPR task. Furthermore, as indicated in Table 7.12 above, no evidence was found of a relationship between performance in the M/C task and performance in the testing phase of the SPR task. In other words, performance in the M/C task did not transfer to performance in the SPR task, suggesting that the two tasks drew on two different knowledge sources.

Further support for the interpretation that performance in the SPR task was reflective of implicit knowledge comes from the nature of implicit versus explicit knowledge. As noted in Section 2.3.3 above, explicit knowledge has been characterised as being available for controlled processing, whereas implicit knowledge is available for more automatic processing (R. Ellis, 2005). Given the nature of the SPR task, it seems unlikely that participants would be able to draw upon explicit knowledge in order to complete this task. This suggests that even the aware participants were not using their explicit knowledge to comb through the sentences carefully in order to identify correct and incorrect instances of case markers. Instead, the aware participants, as did the unaware participants, relied on more automatic knowledge as they processed the sentences for meaning.

At this point, it is worth questioning whether performance in the SPR task is necessarily reflective of implicit knowledge, or if it is simply indicative of more automatic or automatised knowledge. As noted in Section 2.4.2 above, the process of automatisation is often associated with DeKeyser (2001) and with skill acquisition theory. Within this framework, automatisation refers to the process whereby declarative knowledge, such as metalinguistic knowledge, can become functionally equivalent to procedural knowledge as a result of time and practice (DeKeyser, 2007a; 2007b; 2014). Is it possible, then, that the aware participants were able to apply their explicit knowledge automatically? In other words, do the results of the SPR task reflect implicit knowledge that developed independently of explicit knowledge (Paradis, 2004, 2009), or simply explicit knowledge that had become automatised over the course of the experiment (DeKeyser, 2007a, 2007b, 2014)? If the aware participants were utilising explicit knowledge, this might account for the differences in overall reaction times between aware and unaware participants, with aware participants relying on more controlled, explicit processes as they processed the sentences.

Although appealing, such an account is speculative, as the results from the present thesis cannot directly answer the question in the paragraph above. What the data show is that the knowledge does not appear to be explicit and, most importantly, that this knowledge was deployed automatically. Furthermore, as noted above, this knowledge does not appear to be linked to the metalinguistic awareness of the aware participants. However, it is important to note again that awareness here was determined solely via retrospective verbal reports. Given the well-established criticism of the sole reliance on retrospective verbal reports as a measure of awareness (Shanks & St. John, 1994), as well recent criticisms that question their validity in online tasks (Hama & Leow, 2010; Leow & Hama, 2013), it would be wise to exert caution before concluding a dissociation between awareness and the knowledge demonstrated in the SPR task. Future research would undoubtedly benefit from incorporating additional measures, including a concurrent measure of attention, such as eye-tracking (Godfroid et al., 2015). By so doing, these studies would come to a fuller understanding of the conscious status of the knowledge used as part of this task.

In addition, it is important to point out that much of the evidence cited above towards a conclusion of implicit knowledge is, in fact, evidence against a conclusion of explicit knowledge. This problematic line of argument boils down to the knowledge does not appear to be explicit, therefore it is implicit. Such a line of argument appears to be a logical fallacy along the lines of a false dilemma, or a fallacy of exhaustive hypotheses (Searle, 1983). Similar arguments concerning the falsifiability of arguments that learning has proceeded, or that knowledge has been utilised in the absence of awareness, have been raised previously (Baars, 1996; Schmidt, 1995). Simply put, the degree to which evidence that suggests knowledge is not explicit should be interpreted as evidence in favour of implicit knowledge is questionable.

In summary, the results of Experiment 3 point towards the development of both explicit knowledge and automatic knowledge. Explicit knowledge was clearly demonstrated by the aware participants in the M/C task. This result was substantiated by the results from the retrospective verbal reports and subjective measures of awareness. Evidence of automatic knowledge was found through performance in the SPR task, where differences in reaction times to grammatical and ungrammatical sentences served as a measure of learning. Although the fact that automatic knowledge might be argued to be indicative of implicit knowledge (R. Ellis et al., 2009; cf. Leow & Hama, 2013), some caution should be urged before arriving at a similar conclusion here, given the lack of direct evidence available from the present study.

7.6. Limitations and Future Directions

As the results here attest, the use of the SPR task appears to be a promising avenue for future enquiries into the development of implicit and explicit knowledge, particularly when used in conjunction with an offline measure of learning. However, there are a few limitations to the research design of Experiment 3, which should be noted here. The first limitation concerns the design of the SPR task; specifically, whether the segment-by-segment approach was sufficiently sensitive to detect differences in reaction times to grammatical and ungrammatical utterances. SPR tasks can utilise either a segment-by-segment or word-by-word presentation format (see Jegerski, 2014; Keating & Jegerski, 2015; Roberts, 2012a for overviews). One advantage of word-by-word approaches is that such approaches are more sensitive to the detection of differences in reaction times to grammatical and ungrammatical utterances. By contrast, such small differences can potentially be lost in the noise of multi-word segments when using a segment-by-segment approach. Another limitation is that the design of the SPR task did not account for sentence "wrap-up" effects in its analysis of the post-critical region. Wrap-up effects refer to the phenomenon of reading speed tending to slow down at the ultimate segment or word in an SPR task (Jiang, 2012), thus making it more difficult to observe differences in reaction times due to differential amounts of processing. When we consider that the post-critical region was also the ultimate segment in the sentence, it is possible that spillover effects were undetected due to this shortcoming in the research design. Future research could avoid this issue by utilising a word-by-word presentation. Such an approach would result in data that are more robust to the effects of potentially confounding variables.

Another limitation of the present experiment is with regard to statistical power. Statistical power is concerned with the sensitivity of a statistical test when rejecting or accepting the null hypothesis of a particular experiment (Cohen, 1988, 1992a, 1992b; Plonsky, 2013, 2015). A study is underpowered if the sample size is not sufficient to indicate a significant effect with a pre-determined amount of probability (typically 80%, see Cohen, 1988). This can lead to a Type 2 statistical error, such as accepting the null hypothesis when it is, in fact, incorrect. Although other studies investigating implicit learning in SLA are comparable in terms of sample size (Leung & Williams, 2011, 2012), it should be questioned whether the current study (N = 23) was sufficiently powered to uncover the relationships of the fixed effects within the mixed-effects models. The limitations of the sample size of the present study are most apparent with regard to aware (n = 8) and unaware participants (n = 15), in which the number of observations undoubtedly impacted on the accuracy of the statistical analyses used to compare these two groups. For instance, following Cohen's (1988) rules of thumb, the sample size used in Experiment 3 was likely to be able to uncover very large effects between groups. However, a larger sample is needed in order to reveal small- and medium-sized effects. When we consider that research has attested to the difficulties in learning L2 morphology (Long, 2015), it is clear that future research would be strengthened by conducting a power analysis (Cohen, 1988, 1992a, 1992b, 1994; Plonsky, 2014, 2015; Van Voorhis & Morgan, 2007) in the first instance to ascertain the minimum number of participants needed to uncover small to medium effects for the research design in question.

Additionally, larger sample sizes both with regard to the number of participants as well as with regard to the number of items on the assessment measures might allow future research to come to a better understanding of what it is that the participants learned as a result of incidental exposure. In particular, this might illuminate how any learning is affected by the interplay between case, syntax, sentence position, and awareness. For instance, more test items would allow for additional analyses on sub-categories, such as if the learning demonstrated on accusative items was influenced by whether the sentence followed an S-V-O or O-V-S word order. Such data would undoubtedly provide more direct evidence of what is being learned, and might help link the present study with previous research on the processing and learning of case marking within an O-V-S word order (e.g., Culman, Henry, & VanPatten, 2009; Jackson, 2007, 2008).

Finally, future research should also continue to triangulate data from multiple measures. As has been pointed out by many researchers (Chaudron, 2003; Leeser, 2014), there is no perfect instrument, and the use of multiple measures serves to increase the internal validity and generalisability of the results of a study (Purpura et al., 2015). Experiment 3 illustrates the benefits of such an approach, in that the multiple measures of learning and awareness led to a more refined picture of the learning that occurred than would have been gathered from a single instrument used in isolation.

CHAPTER 8

SUMMARY AND CONCLUSION

8.1. Summary of the Thesis

The present thesis reports on the results of five experiments that investigated the learning of an L2 case-marking system under incidental learning conditions, and whether the knowledge resulting from this exposure was implicit or explicit in nature. The stimulus for all five experiments (Pilot Study 1, Pilot Study 2, Experiment 1, Experiment 2 and Experiment 3) was a novel, semi-artificial language that consisted of English phrases with a flexible word order, and which included a single foreign word that had been inflected according to its function within the sentence. This series of experiments implemented a number of methodological improvements to previous research conducted in the same area. For instance, unlike previous research into the role of awareness in the acquisition of L2 case-marking systems, the present thesis did not measure awareness solely using retrospective verbal reports. Instead, all five experiments triangulated data from written and oral retrospective verbal reports in conjunction with subjective measures of awareness. Furthermore, Experiment 3 also incorporated direct and indirect tests of learning in addition to the measures of awareness listed above.

The results of these experiments largely corroborate previous findings (Brooks & Kempe, 2013; Grey et al., 2014; Robinson, 2002, 2005), specifically those that identified the important role of awareness in the learning of L2 case markers. As noted, the previous research found a link between verbalisable rule knowledge, such as awareness at the level of understanding (Schmidt, 1990), and the learning demonstrated by outcome measures in the experiments. This link between metalinguistic knowledge and learning led Grey et al. (2014) to speculate that verbalisable rule knowledge might have been crucial to the learning of the case-marking system in their study. Similarly, Experiment 3 of the present thesis found a similar link between verbalisable rule knowledge and performance in the M/C task.

However, the results of the present thesis also diverged from the studies listed above (Grey et al., 2014; Robinson, 2002, 2005) in that they indicated that some learning can occur in the absence of verbalisable rule knowledge. This result can be seen in the performance in the GJTs in Experiments 1 and 2, where accuracy was not linked to metalinguistic awareness. However, analyses of the subjective measures of awareness within Experiments 1 and 2 indicated that the significant performance of the experimental group coincided with low levels of awareness in the form of noticing. Thus, as in previous research, the results here point towards a relationship between awareness and the learning of L2 case markers; however, the data from this study indicate that lower levels of awareness than suggested by previous research might be sufficient for learning to take place. This result is consistent with research investigating the incidental learning of L2 noun-determiner systems (Leung & Williams, 2012; Williams, 2005) and L2 syntax (Rebuschat & Williams, 2012), which has indicated that participants can acquire some knowledge of L2 grammar without being able to verbalise the rules of the grammatical system in question.

Finally, the results from the SPR task in Experiment 3 indicate that knowledge of L2 case markers acquired as a result of incidental exposure can be deployed automatically. This result is similar to previous research (Leung & Williams, 2011, 2012,

2014) that has incorporated reaction time measures to examine the incidental learning of other areas of L2 grammar. As the design of previous research into the incidental learning of L2 case markers (Grey et al., 2014; Robinson, 2002, 2005) did not include direct measures of learning, this result provides an additional insight into characteristics of the knowledge that can be acquired as a result of incidental exposure.

8.2. Summary of the Individual Experiments

The purpose of the two pilot studies was to establish an initial learning effect of the target case-marking system. The intention was that, once an initial learning effect was established, subsequent studies could manipulate variables within the experiment to examine the effect that these manipulations had on learning, and on the development of implicit and explicit knowledge.

In Pilot Study 1, participants received training on the semi-artificial case-marking system under incidental learning conditions. The stimulus set for this experiment consisted of three different case markers (nominative -*a*, accusative -*u*, and instrumental -*ou*), which were presented within English sentences that followed four different syntactic patterns. In the exposure phase of this experiment, participants listened to the sentence, and then decided which of two pictures presented on the monitor best matched the meaning of the foreign word. The entire exposure phase consisted of 144 sentences, 48 for each of the three morphological categories. Following the exposure phase, participants completed a surprise grammaticality judgement test (GJT), together with subjective measures of awareness. The experiment concluded with retrospective verbal reports in the form of both written and oral interviews. Analyses of the GJT indicated a null result; in other words, participants had not developed knowledge of the target morphological structures. Furthermore, the retrospective verbal reports revealed that participants had not become aware of the morphological rule system. As a result of this null finding, it was not possible to investigate the awareness issue further using subjective measures of awareness.

Pilot Study 2 set out to determine whether decreasing the complexity of the stimulus set and manipulating the training conditions to promote additional noticing (Schmidt, 1990) of the target structures would spark learning of the L2 case-marking system. As in Pilot Study 1, Pilot Study 2 investigated the acquisition of three different morphological markers (nominative -*a*, accusative -*u*, and instrumental -*ou*), but the stimuli were simplified to include only two syntactic patterns (as opposed to four patterns in Pilot Study 1). The training procedure in Pilot Study 2 was also modified in that it included elicited imitations as part of the training task. In the exposure phase, participants first listened to the sentence, then repeated the sentence aloud before indicating which of the two pictures on the monitor best matched the meaning of the foreign word. Despite these modifications, Pilot Study 2 did not result in a significant learning effect. As was the case in Pilot Study 1, the null result for Pilot Study 2 precluded the possibility of examining the data from the subjective measures of awareness. However, data from the retrospective verbal reports indicated that participants had noticed the endings of the foreign words, suggesting that the addition of elicited imitations to the training procedure had resulted in more noticing. Furthermore, participants' comments indicated that they had become aware of the endings because the repetition of the foreign words across the three morphological categories. Thus, the direction forward appeared to be to simplify the training set further, and to manipulate the stimulus set to increase repetition of the stimuli therein.

Experiment 1 followed the same general design as Pilot Studies 1 and 2, but incorporated several changes with regard to both the stimulus set and the training procedure. Firstly, the stimulus set was further simplified to include only two case markers (nominative -*a*, and accusative -*u*). Furthermore, the number of exemplars was reduced from 48 foreign words to only 24. Each foreign word occurred in two unique sentences (one nominative, one accusative) that were each repeated three times across the training set, resulting in a total exposure of 144 sentences. These changes were made to increase the repetition of the exemplars within the training set, while keeping the total amount of exposure across the experiments constant.

The training procedure in Experiment 1 was modified from Pilot Study 2. In this experiment, participants were asked to listen to the sentence, repeat the entire sentence aloud, and to then repeat the foreign word in isolation before judging which of the two pictures on the monitor best matched the meaning of the foreign word. As in the previous pilot studies, the exposure phase here was followed by a surprise GJT, together with subjective measures of awareness. The experiment concluded with retrospective verbal reports, both written and oral. The results of Experiment 1 indicated a significant learning effect for the experimental group when compared to the performance of the control group. Post-hoc tests indicated that performance for accusative items was the driving force behind the overall learning effect demonstrated by the experimental group. Although an analysis of the retrospective verbal reports revealed that this learning effect occurred in the absence of verbalisable rule knowledge, analyses of the subjective measures of awareness provided a more nuanced picture; specifically, that the knowledge acquired was partly implicit in the sense that there was no relationship between confidence and accuracy in the GJT, but also partly explicit in that it was accompanied by low levels of awareness manifested in the form of intuition.

Experiment 2 set out to replicate, validate, and extend the results of Experiment 1. Experiment 2 followed the same basic design as Experiment 1, with the following modifications. The total number of morphological categories reverted to the original three (nominative *-a*, accusative *-u*, instrumental *-ou*). To counteract the potentially detrimental effects that increasing complexity might have on the overall learning effect, the total amount of exposure and the repetition of exemplars was increased in this experiment. Firstly, the total number of exemplars was reduced from 24 to 20. These exemplars were included in three sentences, one for each of the three case markers. Secondly, these sentences were repeated four times each for a total exposure of 240 sentences. All other aspects of the experimental procedure were identical to that of Experiment 1.

The results of Experiment 2 supported the results of Experiment 1. Firstly, the overall learning effect appeared to be driven by performance on accusative markers. Secondly, with regard to awareness, the retrospective verbal reports indicated that this performance took place in the absence of verbalisable rule knowledge, yet was marked by low levels of awareness in the form of noticing the target morphological structures. As in Experiment 1, initial analyses of the subjective measures of awareness indicated that this knowledge was partly implicit and partly explicit in nature. However, further breakdowns of the subjective measures of awareness, specifically when analysing the confidence ratings for accusative items only, found that the knowledge acquired in this experiment was explicit in nature.

The final experiment in this thesis, Experiment 3, set out to investigate the same research questions as the previous experiments, but from a different methodological angle. Experiment 3 incorporated the use of an SPR task as the principle measure of learning, alongside an offline multiple-choice (M/C) sentence completion task as an additional measure. The training set of Experiment 3 followed the general outline of Experiment 1. The stimulus set included two morphological categories (nominative and accusative) and two syntactic patterns. There were 24 foreign words in total, which were included in two sentences (one nominative and one accusative). These sentences were repeated three times each for a total exposure of 144 sentences. The training procedure of the SPR task required participants to click through the sentence, to repeat the sentence silently, and to then answer a comprehension question based on the content of the sentence. In the testing block, participants encountered 48 novel grammatical and ungrammatical sentences. The ungrammatical sentences were generated by replacing the correct case marker with the incorrect case marker. Learning was operationalised as prolonged reaction times to instances of an incorrect case marker. Following the SPR task, participants completed an offline multiple-choice sentence completion task in which they had to indicate which of two options (one nominative and one accusative) best completed a sentence. This was accompanied by subjective measures of awareness. As with the previous experiments in this series, Experiment 3 concluded with both written verbal reports and an oral interview.

The results of Experiment 3 found evidence of learning in both the M/C task and the SPR task. With regard to the M/C task, learning was limited to participants who had developed verbalisable rule knowledge of the morphological rule system, providing strong evidence that their knowledge was explicit in nature. Analyses of the subjective measures of analysis also indicated that the participants had developed explicit knowledge. However, performance in the SPR task, in contrast to performance in the M/C task, was not linked with verbalisable rule knowledge. These results indicate that participants had developed explicit knowledge, as evidenced in the M/C task, as well as more automatic knowledge as measured via the SPR task. However, it should be pointed out that this result could be attributed solely to the fact that only two participants became aware of the case-marking system during the SPR task. As such, this finding should be viewed with some caution until it is supported by additional data.

8.3. Theoretical Implications

The results of the experiments in the present thesis are consistent with previous research investigating the learning of L2 case markers under incidental learning conditions (Grey et al., 2014; Robinson, 2002, 2005); furthermore, they have several implications for our understanding of adult SLA. Firstly, this research has shown that it is possible to develop some knowledge of an L2 case-marking system as a result of incidental exposure, albeit in a semi-artificial language consisting of novel, foreign words nested within a highly established system (i.e., the L1 of the participants). However, as in previous research (Brooks et al., 2011; Brooks & Kempe, 2013; Grey et al., 2014; Robinson, 2002, 2005), this knowledge appeared partial, because performance on outcome measures remained low, particularly with regard to the GJT utilised in

Experiments 1 and 2 and the M/C task in Experiment 3. This suggests there is a limit concerning the aspects of grammar that might be acquired as a result of a limited amount of incidental exposure in the absence of explicit instruction and/or feedback (this point is addressed in more detail below).

In addition, as mentioned above, these results suggest that awareness plays an important, if not vital, role in the process of acquiring L2 case markers. In particular, the data here suggest a link between awareness and the learning of accusative markers attached to novel foreign words. Contrary to previous research that has relied solely on retrospective verbal reports (Brooks & Kempe, 2013; Grey et al., 2014; Robinson, 2002, 2005), the results here indicate that learning can be achieved in the absence of verbalisable rule knowledge; in other words, awareness at the level of understanding (Schmidt, 1990). Evidence for this interpretation can be seen in the results from the GJTs in Experiments 1 and 2, as well as from the data gathered from the SPR task in Experiment 3, in which participants were largely unable to develop metalinguistic understanding of the underlying case-marking system across these experiments. However, all learning in this thesis was accompanied by reports of attention to, and awareness of, the surface features of the morphological forms in the input, namely awareness at the level of noticing (Schmidt, 1990). This result points towards the facilitative role of awareness in the learning process, a result that is consistent with a wealth of empirical research in SLA (Alanen, 1995; Brooks & Kempe, 2013; DeKeyser, 1995; Grey, Williams, & Rebuschat, 2014; Hama & Leow, 2010; Leow, 1997, 2000, 2001; Rebuschat & Williams, 2012; Rosa & Leow, 2004; Rosa & O’Neill, 1999; Robinson, 1997, 2002, 2005; Williams, 2005).

The fact that the learning that took place in these experiments appeared linked to low levels of awareness also concords with theoretical models of attention and awareness in SLA (Robinson, 1995; Schmidt, 1995; Tomlin & Villa, 1994, see Section 2.2 for an overview of these models). In particular, the results here appear to support Schmidt's (1995, 2001) argument that noticing, or conscious registration of the surface features of the stimuli, is facilitative to adult SLA, and is perhaps even necessary under certain conditions, such as in the case of redundant grammatical features (Schmidt, 2001). When we consider the redundant nature of the case-marking system used here, and the fact that learning was only sparked when the training conditions incorporated elicited imitations that necessitated noticing of the target structures, it would appear that noticing might have played an important role in the learning demonstrated in this thesis. Further evidence of the facilitating effect of awareness comes from the M/C task in Experiment 3, in which the performance of participants who developed metalinguistic awareness was significantly higher (~70% accuracy in the M/C task) than was the performance of participants who only demonstrated awareness at the level of noticing (~50% accuracy in the M/C task). This result indicates a relationship between higher levels of awareness and the amount of learning that took place during the experiment. This finding also fits well with Schmidt's (1990) noticing hypothesis, as well as with empirical research that has demonstrated a link between the level of awareness and the amount of learning that takes place (Leow, 1997, 2000; Mackey, 2006; for overviews, see Leow, 2013; Schmidt, 2001, 2010).

Also of interest was the fact that none of the participants who took part in the first four experiments in this series (Pilot Study 1, Pilot Study 2, Experiment 1 and

Experiment 2) managed to crack the code and become aware of the underlying case-marking system. It was only when the modality was changed to a visual presentation in Experiment 3 that some of the participants managed to achieve a metalinguistic understanding of the morphological system. However, it should be pointed out that, although some participants in Experiment 3 came to a conscious understanding of the form-meaning mappings of the case-marking system, the majority of the participants (15/23) failed to do so. This finding echoes the results from previous, similar research. Specifically, this result suggests that the adult learners who took part in this series of experiments appeared largely unable to discover grammatical categories related to inflectional morphology, despite the simplicity and transparency of the case-marking system utilised here.

The fact that learners failed to uncover the rules appears to be in line with the predictions of a number of theoretical positions within SLA. In particular, this finding could be interpreted in light of VanPatten's model of input processing (VanPatten, 2002a), or the shallow structure hypothesis (Clahsen & Felser, 2006), as well as being linked to issues related to perceptual salience, learned attention and blocking (N. Ellis, 2006a). As noted previously, one of the principles of VanPatten's (2004a) model of input processing is related to the primacy of meaning, specifically that learners tend to process input for meaning before they process grammatical forms. Furthermore, VanPatten argued that, when the input contains redundant grammatical forms, as was the case with some of the stimuli used throughout the present thesis, learners are more likely to rely on semantic clues to work out the meaning of a sentence than they are to process these forms (2004a). In a similar vein, these findings could also be seen to support the shallow structure hypothesis (Clahsen & Felser, 2006). As noted in Section 4.3.2, this hypothesis holds that, while L1 speakers fully utilise both lexical-semantic and grammatical cues as they process input, L2 users tend to process input shallowly in that they rely exclusively on lexical-semantic cues to work out the meaning of an utterance. As noted previously, although this hypothesis was initially held to apply only to L2 syntax, more recent research has indicated that L2 learners also process morphology shallowly (Clahsen et al., 2010) Again, given the redundant nature of the stimuli used in this thesis, it appears plausible that participants processed the utterances shallowly and relied on lexical clues, rather than on grammatical morphemes, to comprehend the utterances.

Finally, N. Ellis (2006a, 2006b) has postulated that attention is learned and selective, and, furthermore, that the salience of input is shaped by both the learner's L1 and the features of the input itself. In this sense, attention can be seen as an aspect of language that can potentially transfer between the learners' L1 and their L2. N. Ellis argued that the salience of cues is influenced by the redundancy of cues. If, for example, we take the sentence "I walked to the store last week," the temporal meaning is marked both lexically (through the adverbial phrase *last week*) and grammatically, through the -*ed* morphological marker. According to N. Ellis (2006a), in cases of redundant cues, learners might never notice the lower salience cues, particularly when the more obvious cues suffice for the communicative task at hand. In this sense, the cue with the higher degree of salience overshadows, and possibly blocks, the lower salience cue. This has several ramifications for the present thesis. Firstly, when we consider that the L1 of the participants in the present thesis (English) does not include a case-marking system, it seems possible that the participants were not naturally tuned to attend to the word endings as a cue to the meaning of the utterance. Secondly, as the meaning carried by the case markers was redundant, this makes it less likely that the learners would necessarily attend to these forms, particularly as the case marker would represent the less salient cue in N. Ellis' (2006a, 2006b) model.

In summary, VanPatten's model of input processing, the shallow structure hypothesis, as well as issues related to perceptual salience, can be seen to provide an explanation for the limited amount of learning displayed in the experiments in this thesis. In short, it would appear that the participants relied on the semantic information conveyed by the lexis of the sentence, as well as on the contextual information provided by the pictures (in Experiments 1 and 2), in order to comprehend the utterances. It was only when the training conditions necessitated that the participants attend to the surface features of the input, in the form of elicited imitations, that some learning of the morphological system took place.

To shift the focus towards the development of implicit and explicit knowledge, the results of Experiments 1 and 2 indicated that participants developed knowledge that was characterised by low levels of awareness. In other words, the participants were aware that they had learned something (Dienes, 2004), but this awareness remained below the threshold of verbalisation. The results of Experiment 3 were largely similar to those of the previous two experiments, with the exception that some of the participants in Experiment 3 developed verbalisable rule knowledge of the target case marking system. This result is interesting for a number of reasons. Firstly, it underscores the fact that incidental learning conditions are no guarantee that learning will take place incidentally, nor that the knowledge acquired as a result of incidental exposure will necessarily be implicit in nature. In fact, analyses of the retrospective verbal reports in Experiment 3 indicated that a number of the participants admitted to engaging in rule-search behaviour during the experiment, with the majority (6/8) doing so during the M/C task. Such a result mirrors findings from previous research that reported similar behaviour among participants under incidental conditions (DeKeyser, 1995, 1997; Robinson, 1997, 2002, 2005). However, this finding might have been influenced by the nature of the M/C task, which required participants to choose between two forms of a foreign word, one with the accusative marker and one with the nominative marker. This dichotomous choice is likely to have encouraged rule-search behaviour in that it is clear that the case marking is linked to the correct answer.To address this issue, future research might examine the performance of participants across the testing phase of the experiment to determine if there is a difference in performance between, for example, the first quarter versus the last quarter of the M/C task. An increase in performance would provide evidence that participants had become aware during the testing phase and not at a previous point in the experiment.

It is also interesting to consider how the above results might be interpreted with regard to the interface between implicit and explicit knowledge (see Section 2.4 for a discussion). For instance, the fact that participants appeared to engage in rule-search behaviour in Experiment 3 could be offered as a simple explanation for the reason that participants developed explicit knowledge. In other words, the use of explicit learning strategies resulted in explicit knowledge. Such a result might be seen to support non-interface positions (Krashen, 1982) in which explicit knowledge and implicit knowledge arise separately out of isolated processes. However, this might not necessarily be the case. For instance, recent studies have demonstrated that learners develop explicit knowledge rapidly as a result of incidental learning conditions, even in the absence of reported rule-search behaviour (Hamrick, 2013).

Thus, this result, as well as the overall results from the present thesis, could support N. Ellis' (2005) weak interface position. As part of N. Ellis' model, mental representations are first seeded within explicit memory before developing into abstract "trees," or implicit knowledge (see Section 2.4.3 for a discussion). This process occurs over a long period as a result of repeated exposure to patterns within the input. Following this model, the fact that the learning demonstrated in the experiments within this thesis was marked by conscious awareness could be explained by the theoretical account above; specifically, that more explicit knowledge is quick to develop, whereas implicit knowledge develops slowly over a longer period as implicit mechanisms work to abstract grammatical regularities contained in the input (N. Ellis, 2005; Robinson, 1996, 1997).[40]

However, it is important to note that such an account is speculative, as this thesis provides no direct evidence of the interface between implicit and explicit knowledge, nor of how these types of knowledge develop over time. Although some research has investigated the time course of implicit and explicit knowledge with regard to the learning of artificial grammars (Ziori & Dienes, 2012), to my knowledge, no SLA studies have yet done so. By tracing how implicit and explicit knowledge develop over time, future research might better inform the debate surrounding the interface between implicit and explicit knowledge in SLA.

[40] However, such data could be interpreted to support the declarative / procedural model of learning (see DeKeyser, 2009; Ullman, 2004 for discussions).

8.4. Methodological Implications

The results of this thesis have a number of methodological implications. Firstly, the results of Experiments 1 and 2 underscore the fact that relying on retrospective verbal reports as the sole measure of awareness is an inadequate means of measuring the nature of knowledge acquired within a study (Rebuschat, 2013; Shanks & St. John, 1996). Furthermore, it provides additional evidence of the usefulness of subjective measures of awareness in identifying low levels of awareness, which might otherwise be missed by retrospective verbal reports.

However, it should be pointed out that Experiment 2 also suggested a limitation of the zero-correlation criterion as a measure of implicit knowledge (Dienes & Scott, 2005). As noted, learning in Experiment 2 was limited to accusative items. When the entire dataset was analysed (both nominative and accusative items), analyses revealed no significant relationship between confidence and accuracy in the GJT. Following the zero-correlation criterion, this would provide evidence that the knowledge acquired was implicit in nature. However, when the accusative items in the GJT were examined in isolation, the results indicated a statistically significant relationship between confidence and accuracy, indicating that the knowledge acquired was, in fact, explicit in nature. In summary, it would appear that confidence ratings can be influenced by performance on a sub-component of a test, just as the overall learning effect might be driven or masked by performance on a particular type of test item. In this sense, future research might benefit from analysing confidence rating data across all sub-components of a test. Such an undertaking could lead to a more refined and valid understanding of the conscious status of knowledge being utilised in a particular assessment.

In addition, with regard to incorporating multiple measures of learning, the results of Experiment 3 underscore the benefits of including both direct and indirect measures of learning within a single experiment. For instance, if Experiment 3 had relied solely on the M/C task, the results would have only provided evidence that the knowledge acquired was strongly linked with verbalisable rule knowledge. This result would have been interpreted as providing evidence that the incidental exposure conditions led to purely explicit knowledge. Conversely, if this experiment had utilised only the SPR task, the results would have indicated no link between verbalisable rule knowledge and learning. This result would have been interpreted to indicate that the incidental learning conditions resulted in implicit knowledge. Instead, the use of both of these measures allowed for a more nuanced view of the learning that had taken place in this study: specifically, that the incidental exposure conditions appeared to result in two types of knowledge, or in a single type of knowledge that was utilised in different ways. Such a finding highlights the benefits of triangulating data from multiple measures of awareness and multiple measures of learning (Purpura et al., 2015). Furthermore, this result provides additional support for the use of both direct and indirect tests when investigating implicit and explicit phenomena in SLA (R. Ellis et al., 2009; Leung & Williams, 2011; Norris & Ortega, 2000; Reingold & Merikle, 1988).

Finally, when reflecting critically on the measures of awareness included in the present study, a limitation of Pilot Study 1, Pilot Study 2, Experiment 1, and Experiment 2 lies in the lack of information about the strategies adopted by participants when dealing with the stimuli in the exposure phase. Incidental learning conditions as operationalised here provide no guarantee that learners actually learn incidentally; that is, without intention. This issue was addressed in Experiment 3 by modifying the retrospective verbal reports to include a question that asked whether participants were actively searching for rules during the experiment. As in previous research (Robinson, 2002, 2005), responses to this question indicated that some participants in Experiment 3 engaged in rule-search behaviour during the experiment, despite the fact that the experimental condition was designed to create opportunities for incidental learning. Future studies, similarly to earlier research in this area (DeKeyser, 1995; Robinson 1997, 2002, 2005), should continue to include retrospective questions asking participants if they were actively involved in searching for rules. This information could increase the construct validity of incidental learning studies. In particular, it could reveal whether the intended incidental learning conditions succeeded in orienting learners' attention towards the meaning of the stimuli and in preventing them from utilising explicit learning strategies.

8.5. Limitations and Future Directions

Despite the positive results listed above, this thesis leaves a number of questions connected with the development of implicit and explicit knowledge unanswered. Firstly, Experiments 1 and 2 witnessed the development of knowledge that was made manifest in the form of intuition judgements in a GJT.Although some have argued that such intuition judgements are reflective of implicit knowledge (Dienes, 2008), others have argued against the possibility of participants gaining native-speaker-like intuition after a short exposure period (Hama & Leow, 2010; Serafini, 2013). Implicit in the latter argument is that the construct of intuition, as indicated via source attribution data, is not equivalent to the intuition that native speakers have of their L1. In other words, there is an assumption that similar intuition judgements by L1 users would not be accompanied by low levels of awareness. To my knowledge, however, no research to date has set out to examine if this is necessarily the case. A further limitation of the subjective measures of awareness was highlighted in Experiment 2, in which the data cast some doubt on the reliability of the zero-correlation criterion as criterion for implicit knowledge. To my knowledge, no research to date has thoroughly validated subjective measures of awareness as reliable and valid tools towards the measurement of implicit and explicit knowledge. If one goal of SLA is to foster the development of implicit knowledge, then validation studies of these instruments is clearly warranted.

Also, as has been noted by many others (Hulstijn, 1997), caution must be exercised when attempting to extrapolate laboratory studies to naturalistic language acquisition, particularly when these experiments utilise artificial grammars, the generalisability of which to natural languages has long been challenged (Schmidt, 1994b). As noted previously (please refer to Section 3.4.2), the use of artificial languages stems from the quantitative tradition that stresses "reliability over validity and control over context" (Hulstijn et al., 2014, p. 365). In this sense, the advantages of artificial languages are clear–they allow the experimenter complete control over the input present in the experiment, and allow the experiment to control effectively for other confounding variables, such as prior knowledge on the part of the participants.

However, the limitations of artificial languages, such as the system employed in the present study, are also apparent. For example, as noted in Section 5.1, some authors (e.g., Rogers, Revesz, & Rebuschat, 2015) have argued for the need to demonstrate that a semi-artificial system is learnable within a particular context as part of the experimental piloting process. This argument is predicated on the many unknowns when working with a novel semi-artificial system, such as whether the system might be too simple or too complex for a particular learning context. To illustrate, Pilot Studies 1 and 2 found no learning effect of the target morphological structures. In isolation, the results from these two experiments might be interpreted to indicate that that no implicit learning, or no development of implicit knowledge, is possible within that particular context. However, it is important to note that these pilot studies indicated no learning had taken place, neither implicit nor explicit. To be able to say with any certainly that implicit learning is not possible within a particular context, one must demonstrate that explicit learning is possible. Otherwise, one must simply conclude that no learning is possible. Thus, validation of the learnability of semi-artificial languages appears warranted, in particular when used to investigate issues related to attention and awareness in SLA.

Further limitations of semi-artificial languages include the fact that these languages represent highly simplified versions of natural language systems, and often completely ignore important aspects of language, such as pragmatics. In addition, it is important to stress that there is no direct evidence that the cognitive processes involved in processing and learning the semi-artificial system employed in this thesis necessarily reflect that of processing and learning a natural language. For example, in the present thesis, the presence of a single foreign word embedded within familiar lexis might have resulted in the participants allocating an unnatural amount of attention to the foreign word. Such behavior would call into question whether the findings reported here, in particular findings related to the role of awareness in SLA, necessarily generalise to the learning of natural languages. Future research would undoubtedly be strengthened by developing a process to validate the semi-artificial system in question, or by using stimuli that reflect the complexities inherent within natural languages more closely (for example, a miniature language system such as that which was used in Brooks & Kempe, 2013).

In addition to the more general limitations related to the use of semi-artificial languages in SLA research, there are also a number of limitations specific to the semi-artificial system utilised within this thesis. In particular, although care was taken in the creation of the semi-artificial system, there are a number of variables that were not controlled for and, as such, might have impacted on the results of the study. For instance, the stimuli included sentences that contained both redundant and essential case markers. In addition, the animacy of the foreign noun as well as the use of pronouns (e.g., “he”) and names (e.g., “Patrick”) were not effectively balanced across the training and testing sets. The internal validity of future research would undoubtedly be strengthened by modifying the stimuli to better control for these potentially confounding variables.

Despite these reservations, it should be noted that recent research has pointed towards similarities between the cognitive processes that underpin the processing of natural languages and the processing of artificial grammars (Folia, Uddén, De Vries, Forkstam, & Petersson, 2010; Petersson, Folia, & Hagoort, 2012), as well as a relationship between the learning of artificial grammars and the learning of an L2 (Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, in press; cf. Robinson, 2005). Finally, when we consider the amount of research that has utilised artificial or semi-artificial languages that has undoubtedly contributed to our understanding of SLA (de Graaf, 1997; DeKeyser, 1995, 1997; Grey et al., 2014; Hama & Leow, 2010; Leung & Williams, 2011, 2012, 2014; Rebuschat & Williams, 2012; Robinson, 1997, 2010;

Williams, 2005[41]), it is hoped that the semi-artificial language used here can also be a valid instrument in contributing to our knowledge of language learning processes.

To shift the focus towards possible future directions for this line of research, there are a number of issues related to implicit processes within SLA that have yet to be determined. For instance, there is ongoing debate regarding the degree to which individual difference variables affect implicit and explicit processes in SLA (Paradis, 2004, 2009). A number of cognitive individual difference variables have been implicated within this debate. These include both phonological short-term memory and/or working memory (N. Ellis, 2005; Paradis, 2009; Roehr, 2008; Williams, 2012), personality and learning styles (Grey, Williams, & Rebuschat, 2015; Kaufman et al., 2011), information processing cognitive style (Granena, in press), and sequence learning ability (Granena, 2013; Kaufman et al., 2010). Although some research has attempted to address the role that some of these different variables might play within implicit and explicit learning processes, the research to date has been limited, and has produced conflicting results (see, for example, the relationship between working memory and incidental learning in Tagarelli et al., 2015, versus Robinson, 2002). It is clear that further research in this area is necessary in order to develop a fuller and more valid understanding of the cognitive variables that underpin the development of implicit and/or explicit knowledge.

Finally, the development of automatic knowledge appears to be an area that most strongly warrants further research and is the avenue that my future research will follow. The development of automatic knowledge is interesting for a number of reasons. On a theoretical level, such research might help inform the debate surrounding the

[41] See also Gómez (2007) for a review of the use of artificial grammars in L1 research.

development of implicit and explicit knowledge in that it might lead to a better understanding of the defining characteristics of these types of knowledge within the field of SLA. On a practical level, the develop of automatic knowledge has immediate implications for the L2 classroom. To elaborate, it can be assumed that the goal of the average language learner is to be able to use his or her L2 fluently, intuitively, naturally, and efficiently in a variety of situations. In this sense, whether the knowledge used is accompanied by awareness might not be important to the average language learner. Instead, he or she is concerned with whether this knowledge can be deployed automatically and accurately. Thus, one direction forward might be to examine the development of knowledge that can be used spontaneously in both comprehension and production (Keenan & MacWhinney, 1987). Researchers could examine the degree to which various learning conditions, such as incidental versus intentional exposure, lead to automatic knowledge and, furthermore, could investigate the role of awareness within this process. Such an understanding would not only benefit the field on a theoretical level, but could also provide results with more immediate implications for the L2 classroom.

As can be seen, the results of this thesis have raised more questions than they have answered. It is hoped that future research will continue to improve upon the methodology employed here. Such an undertaking might allow for a more complete understanding of how implicit and explicit knowledge develop, and how these two types of knowledge might best be measured. Despite the limitations of the experiments in this thesis, and the remaining unanswered questions, it is important to highlight that this thesis can be seen as part of an ongoing narrative. In other words, the completion of this thesis does not represent a conclusion of this line of investigation but, instead, a second starting point. The goal of my future research is to attempt to answer some of the questions outlined in the paragraphs above. It is hoped that such an undertaking will contribute to the debate regarding the development of implicit and explicit knowledge, and will allow for a fuller understanding of the cognitive processes that underlie adult SLA.

REFERENCE LIST

Abbuhl, R. (2011). Why, when and how to replicate research. In A. Mackey & S.M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 296-312). London: Basil Blackwell.

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

Allen, R., & Reber, A.S. (1980). Very long term memory for tacit knowledge. *Cognition*, *8*,175‒185.

Amato, M.S., & MacDonald, M.C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*, 143–148.

Andersen, R.W. (1984). The one to one principle of interlanguage construction. *Language Learning*, *34*, 77-95.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

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

Anderson, J.R. (1990). *Cognitive psychology and its implications*. New York: W.H. Freeman.

Anderson, J.R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, *105*, 165-180.

Anderson, J.R. (2000). *Learning and memory*. New York: Wiley.

Anderson, J.R., & Schunn, C.D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser, (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Volume 5, pp. 1-34). Mahwah, N.J.: Lawrence Erlbaum Associates.

Andringa, S., & Curcic, M. (September, 2014). *How metalinguistic information affects online L2 processing: Evidence from differential object marking (DOM) acquisition*. Paper presented at the 24th annual conference of the European Second Language Association (EUROSLA). York, UK.

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition, 37*, 237-268.

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, *37*, 185-196.

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

Baars, B.J. (1997a). *In the theater of consciousness: The workspace of the mind*. Oxford: Oxford University Press.

Baars, B.J. (1997b). In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness* Studies, *4*, 292-309.

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Baayen, R.H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12-28.

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

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

Baddeley, A.D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A.D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-89). New York: Academic Press.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bardovi-Harlig, K. (2007). One functional approach to second language acquisition: The concept-oriented approach. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 57-75). Mahwah, N.J.: Lawrence Erlbaum Associates.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bartek, B., Lewis, R.L., Vasishth, S., & Smith, M.R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1178-1198.

Bartoń, K. (2013). MuMIn: multi-model inference. R package version, 1(5).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN. R-project. org/package= lme4.

Bauernfeind, R.H. (1968). The need for replication in educational research. *The Phi Delta*

Kappan, *50*, 126-128.

Bayne, T., Cleeremans, A., & Wilken, P. (Eds.). (2009). *The Oxford companion to consciousness*. Oxford: Oxford University Press.

Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). *Regression diagnostics. Identifying influential data and sources of collinearity*. New York: Wiley.

Benati, A. (2013). The Input Processing Theory. In M. Pilar Garcia Mayo, G. Mangado, & A. Martinez (Eds.), *Contemporary approaches to second language acquisition* (pp. 93-110). Amsterdam: John Benjamins.

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

Bialystok, E. (1982). On the relationship between knowing and using forms. *Applied Linguistics*, *3*,181-206.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, *57*, 498-510.

Bialystok, E. (1990). The competence of processing: Classifying theories of second language acquisition. *TESOL Quarterly*, *24*, 635-648.

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology*, *28*, 654-664.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*, 157-168.

Bialystok, E., & Craik, F.I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*, 19-23.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*, 635-650.

Blackmore, S. (2010). *Consciousness: An introduction*. London: Hodder Education.

Block, N. (1996). How can we find the neural correlate of consciousness? *Trends in Neurosciences*, *19*, 456-459.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioural and Brain Sciences*, *30*, 481-548.

Bodenhausen, G.V. (1990). Stereotypes as judgemental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*, 319-322.

Bowles, M.A. (2008). Task type and reactivity of verbal reports in SLA: A first look at an L2 task other than reading. *Studies in Second Language Acquisition*, *30*, 359-387.

Bowles, M.A. (2010a). *The think-aloud controversy in language acquisition research*. New York: Routledge.

Bowles, M.A. (2010b). Concurrent verbal reports in second language research. *Annual Review of Applied Linguistics*, *30*, 111-127.

Bowles, M.A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*, 247-271.

Bowles, M.A., & Leow, R.P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*, *27*, 415-440.

Box, G.E., & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*, 211-252.

Braine, M.D.S. (1963). On learning the grammatical order of words. *Psychological Review*, *70*, 323-348.

Braine, M.D.S. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, *72*, 532-540.

Brody, N. (1989). Unconscious learning of rules: Comment on Reber's analysis of implicit learning. *Journal of Experimental Psychology: General*, *118*, 236-238.

Brooks, L.R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 170-211). Hillsdale, N.J.: Erlbaum.

Brooks, P.J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory and Cognition*, *41*, 281-296.

Brooks, P.J., Kempe, V., & Donachie, A. (2011). Second language learning benefits from similarity in word endings: Evidence from Russian. *Language Learning*, *61*, 1142-1172.

Brooks, P.J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*, 185-209.

Brown, D. (2011). Morphological typology. In J.J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp.487-503). Oxford: Oxford University Press.

Brown, J.D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.

Brown, J.D. (2015). Why bother learning advanced quantitative methods in L2 research? In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 9-20). New York: Routledge.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10,* 425-455.

Bybee, J., & Thompson, S. (2000). Three frequency effects in syntax. *Berkeley Linguistic Society*, *23*, 65-85

Cadierno, T. (2000). The acquisition of Spanish grammatical aspect by Danish advanced language learners. *Spanish Applied Linguistics*, *4*, 1-53.

Carlson, R.A. (2003). Skill learning. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol 4, pp. 36-42). London: Macmillan.

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

Carroll, S.E. (2004). Some comments on input processing and processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 293-310). Mahwah, N.J.: Lawrence Erlbaum Associates.

Carroll, S.E. (2007). Autonomous induction theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 155-200). Mahwah, N.J.: Lawrence Erlbaum Associates.

Cattell, R.B., & Cattell, H.E.P. (1973). *Measuring intelligence with the Culture Fair Tests*. Champaign, IL: Institute for Personality and Ability Testing.

Cedrus Corporation. (2012). SuperLab (Version 4.5). San Pedro, CA.

Chaudron, C. (1983). Research on metalinguistic judgements: A review of theory, method and results. *Language Learning*, *33,* 343-377.

Chaudron, C. (2003). Data collection in SLA research. In C.J. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 762-828). Oxford: Blackwell.

Cheesman, J., & Merikle, P.M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology*, *40*, 343-367.

Clahsen, H., Eisenbeiss, S., & Penke, M. (1996). Lexical learning in early syntactic development. In H. Clahsen (Ed.), *Generative perspectives on language acquisition: Empirical findings, theoretical considerations and crosslinguistic comparisons* (pp. 129-159). Amsterdam: John Benjamins.

Clahsen, H., Felser, C., Neubauer, K., Sato, M. & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, *60*, 21-43.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3-42.

Clahsen, H., & Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua*, *120*, 2627-2637.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, *12*, 335-359.

Clark, E.V. (2001). Morphology in language acquisition. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology* (pp. 374-389). Oxford: Blackwell.

Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cognitive Science*, *38*, 1286-1315.

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* (pp. 1-40). New York: Psychology Press.

Cleeremans, A., & McClelland, J.L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235-253.

Cobb, T. (2012). Web Vocabprofile [accessed 14 November 2012 from http://www.lextutor.ca/vp/], an adaptation of Heatley & Nation's (1994) Range.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213-220.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, N.J: Lawrence Erlbaum Associates.

Cohen, J. (1992a). A power primer. *Psychological bulletin*, *112*, 155-159.

Cohen, J. (1992b). Statistical power analysis. *Current Directions in Psychological Science*, *1*, 98-101.

Cohen, J. (1994). The Earth is round (p < .05). *American Psychologist*, *49*, 997-1003.

Coleman, E.B. (1964). Generalizing to a language population. *Psychological Reports*, *14*, 219-226.

Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly*, *45*, 106-133.

Comrie, B. (Ed.) (2009). *The world's major languages* (2nd Ed.). New York: Routledge.

Cook, V.J. (1988). Language learners' extrapolation of word order in micro-artificial languages. *Language Learning*, *38*, 497-529.

Corder, S.P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*, 161-169.

Cox, J.G., & Sanz, C. (2015). Deconstructing PI for the ages: Explicit instruction vs. practice in young and older bilinguals. *International Review of Applied Linguistics in Language Teaching*, *53*, 225-248.

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

Creel, S.C., Newport, E.L., & Aslin, R.N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119-1130.

Culman, H., Henry, N., & VanPatten, B. (2009). The role of explicit information in instructed SLA: An on-line study with processing instruction and German accusative case inflections. *Die Unterrichtspraxis/Teaching German*, *42*, 19–31

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*, 369-382.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117-139.

Dabrowska, E., & Szczerbinski, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language*, *33*, 559-597.

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

de Bray, R.G.A. (1980). *Guide to the West Slavonic languages*. Columbus: Slavica Publishers.

de Graaf, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249-297.

DeKeyser, R.M. (1994). Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly*, *28*, 188-94.

DeKeyser, R.M. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition*, *17*, 379-410.

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

DeKeyser, R.M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom language acquisition* (pp. 42-63). New York: Cambridge University Press.

DeKeyser, R.M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-51). New York: Cambridge University Press.

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

DeKeyser, R.M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1-25.

DeKeyser, R.M. (2007a). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97-113). Mahwah, N.J.: Lawrence Erlbaum Associates.

DeKeyser, R.M. (2007b). Situating the concept of practice. In R.M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). Cambridge: Cambridge University Press.

DeKeyser, R.M. (2007c). The future of practice. In R.M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 287-304). Cambridge: Cambridge University Press.

DeKeyser, R.M. (2009). Cognitive–psychological processes in second language learning. In M. Long & C. Doughty (Eds.), *Handbook of second language teaching* (pp. 119-138). Oxford: Blackwell.

DeKeyser, R.M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed., pp. 92-112). New York: Routledge.

DeKeyser, R.M., & Juffs, A. (2005). Cognitive considerations in L2 learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 437-454). Mahwah, N.J.: Lawrence Erlbaum Associates.

DeKeyser, R.M., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "Processing instruction: An update". *Language Learning*, *52*, 805-823.

Descartes, R. (2010). *The principles of philosophy*. (Trans: J. Veitch). Original work published 1644. Whitefish, Mont: Kessinger Pub.

de Zeeuw, M., Verhoeven, L., & Schreuder, R. (2012). Morphological family size effects in young first and second language learners: Evidence of cross-language semantic activation in visual word recognition. *Language Learning*, *62*, 68-92.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies*, *11*, 25-45.

Dienes, Z. (2008a). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York: Palgrave Macmillan.

Dienes, Z. (2008b). Subjective measures of unconscious knowledge. *Progress in Brain Research, 168*, 49-64.

Dienes, Z. (2010). Measuring the conscious status of knowledge. In L.A.P. Miranda & A.I. Madariaga (Eds.), *Advances in cognitive science: Learning, evolution and social action* (pp. 113-128). Bilbao, Spain: University of the Basque Country Press.

Dienes, Z. (2013). Conscious versus unconscious learning of structure. In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 337-364). Berlin: De Gruyter.

Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains? How implicit and how abstract? In D. Berry (Ed.), *How implicit is implicit learning?* (pp 107-123). Oxford: Oxford University Press.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1322-1338.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin and Review*, *4*, 3-23.

Dienes, Z., Broadbent, D.E., & Berry, D.C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 875-882.

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

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Oxon: Routledge.

Doughty, C.J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C.J. Doughty & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 256-310). Malden: Blackwell.

Dressler, W.U. (2012). On the acquisition of inflectional morphology: Introduction. *Morphology*, *22*, 1-8.

Dulany, D.E., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgement: How concrete and how abstract? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *113*, 541-555.

Dussias, P.E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition*, *25*, 529-557.

Dussias, P.E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149-166.

Dussias, P.E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101-116.

Earp, B.D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 1-11.

Eckman, F., Bell, L., & Nelson, D. (1988). On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics*, *9*, 1-20.

Eisenbeiss, S., Bartke, S., & Clahsen, H. (2006). Structural and lexical case in child German: Evidence from language-impaired and typically developing children. *Language Acquisition*, *13*, 3-32.

Eisenbeiss, S., Narasimhan, B., Voeikova, M. (2008). The acquisition of case. In A. Malchukov & A. Spencer (Eds.), *The Oxford handbook of case* (pp. 369-383). Oxford: Oxford University Press.

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

Ellis, N.C. (Ed.) (1994a). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N.C. (1994b). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, *11*, 37-56.

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

Ellis, N.C. (1999). Cognitive approaches to SLA. *Annual Review of Applied Linguistics*, *19*, 22-42.

Ellis, N.C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Ellis, N.C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143-188.

Ellis, N.C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 33-68). 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. (2006a). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 164-194.

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

Ellis, N.C. (2007a). Dynamic systems theory and SLA: The wood and the trees. *Bilingualism: Language and Cognition*, *10*, 23-25.

Ellis, N.C. (2007b). Implicit and explicit knowledge about language. In J. Cenoz & N.H. Hornberger (Eds.), *Encyclopedia of Language and Education*, *Second Edition, Volume 6: Knowledge about Language* (pp. 119-132). New York: Springer.

Ellis, N.C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *Modern Language Journal*, *92*, 232-49.

Ellis, N.C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R.P. Leow (Eds.) *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA & bilingualism* (pp. 35-47). Washington, DC: Georgetown University Press.

Ellis, N.C. (2015b). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 1-24). Amsterdam: John Benjamins.

Ellis, N.C., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function: Introduction to the special issue. *The Modern Language Journal, 93*, 329-335.

Ellis, N.C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics. Introduction to the special issue. *Applied Linguistics, 27*, 558-589.

Ellis, N.C., & Sagarra, N. (2010a). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, *32*, 553-580.

Ellis, N.C., & Sagarra, N. (2010b). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning*, *60*, 85-108.

Ellis, N.C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, *33*, 589-624.

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, N.C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75-93). New York: Routledge.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly*, *27*, 91-113.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 223-236). London: Academic Press.

Ellis, R. (2004).The definition and measurement of L2 explicit knowledge. *Language Learning*, *52*, 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141-171.

Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2015). Form-focused instruction and the measurement of implicit and explicit knowledge. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 417-442). Amsterdam: John Benjamins.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

Emonds, J.E. (1985). *A unified theory of syntactic categories*. Dordrecht: Foris.

Engel, R., & Schutt, R. (2010). *Fundamentals of social work research*. Los Angeles: Sage.

Erlam, R. (2013). Effects of instruction on learners' acquisition of metalinguistic knowledge. In K. Roehr & G.A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 71-94). London: Bloomsbury.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P.C.M. (in press). The relationship between artificial and second language learning. *Cognitive Science*.

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena et al. (Eds.), *Selected proceedings of the 2010 second language research forum: Reconsidering SLA research, dimensions, and directions* (pp. 18-28). Somerville, MA: Cascadilla Proceedings Project.

Felser, C. (2012). Language processing evidence for linguistic structure. *Iberia*, *4*, 1-22.

Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research*, *38*, 305-319.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453-489.

Fernández, E.M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins.

Ferreira, F., & Henderson, J.M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 555-568.

Folia, V., Uddén, J., De Vries, M., Forkstam, C., & Petersson, K.M. (2010). Artificial language learning in adults and children. *Language Learning*, *60*, 188-220.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. *Advances in Psychology*, *134*, 217-236.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, *50*, 119-148.

Frenck-Mestre, C., & Pynte, J. (2000). Resolving syntactic ambiguities: Cross-linguistic differences? In M. de Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspectives on language processing* (pp. 119-148). Amsterdam: Springer.

Gagarina, N., & Voeikova, M. (2009). Acquisition of case and number in Russian. In U. Stepany & M.D. Voeikova (Eds.), *Development of nominal inflection in first language acquisition* (pp. 179-216). Berlin: Mouton De Gruyter.

Gagné, C.L., & Spalding, T.L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, *60*, 20-35.

Gass, S.M. (1979). Language transfer and universal grammatical relations. *Language Learning*, *29*, 327-344.

Gass, S.M. (1982). From theory to practice. In M. Hines & W. Rutherford (Eds.), *On TESOL '81: Selected Papers of the Fifteenth Annual Conference of Teachers of English to Speakers of Other Languages* (pp. 129-139). Washington, DC: TESOL.

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

Gass, S.M., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Mivashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Boston: MIT Press.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgements measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269-297.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S.N. Frota & J.K. Yoshioka (Eds.), *Noticing and second language* *acquisition: Studies in honor of Richard Schmidt* (pp. 183-205). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Godfroid, A., & Winke, P. M. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 323-348). Amsterdam: John Benjamins.

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. (2007). Statistical learning in infant language development. In M. Gareth Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 601-616). Oxford: Oxford University Press.

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

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.

Gor, K. (2010). Introduction. Beyond the obvious: Do second language learners process inflectional morphology? *Language Learning*, *60*, 1-20.

Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, *28*, 1065-1091.

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665-703.

Granena, G. (in press). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*.

Gregg, K. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics*, *5*, 79-100.

Grey, S., Williams, J.N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*, 1-34.

Grey, S., Williams, J.N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, *38*, 44-53.

Gries, S.T. (2013). *Statistics for linguistics with R: A practical introduction*. Berlin: De Gruyter.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology*, *56*, 1213–1236.

Hahne, A., Mueller, J. & Clahsen, H. (2006). Morphological processing in a second language: Behavioural and event-related potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, *18*, 121-134.

Hama, M., & Leow, R.P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition*, *32*, 465-491.

Hamrick, P. (2013). *Development of conscious knowledge during early incidental learning of L2 syntax* (Unpublished doctoral dissertation). Georgetown University: Washington, D.C.

Hamrick, P. & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365-382). Berlin: Mouton De Gruyter.

Hamrick, P. & Rebuschat, P. (2013). Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L. Amoroso (Eds.), *Measured language: Quantitative approaches to acquisition, assessment, processing, and variation* (pp. 125-139). Georgetown University Press: Washington, D.C.

Han, Y. (2000). Grammaticality judgment tests: How reliable and valid are they? *Applied Language Learning*, *11*, 177-204.

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

Han, Z-H., & Finneran, R. (2014). Re-engaging the interface debate: strong, weak, none, or all? *International Journal of Applied Linguistics*, *24*, 370-389.

Han, Z-H., & Sun, Y.A. (2014). First exposure: a replication of Han and Peverly (2007). In Z-H. Han & R. Rast (Eds.), *First exposure to a second language: Learners' initial input processing* (pp. 7-39). Cambridge: Cambridge University Press.

Hanson, B. (2011). Slavonic languages. In B. Kortmann & J. van der Auwera (Eds.), *The languages and linguistics of Europe* (pp. 97-123). Berlin: de Gruyter.

Harrell, F.E., & Dupont, C. (2015). Hmisc: Harrell miscellaneous. R package version 3.9-3. Available at http://CR AN. R-project. org/package= Hmisc.

Harrington, M. (2004). Commentary: Input processing as a theory of processing input. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 79-92). Mahwah, N.J.: Lawrence Erlbaum Associates.

Harrington, M. (2010). Computation models of second language sentence processing. In W. Grabe, M. Swain, & G.R. Tucker (Eds.), *The Oxford handbook of applied linguistics* (pp. 189-203). Oxford: Oxford University Press.

Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences, 33*, 61-83.

Hinz, J., Krause, C., Rast, R., Shoemaker, E. & Watorek, M. (2013). Initial processing of morphological marking in nonnative language acquisition: Evidence from French and German learners of Polish. In L. Roberts, A. Ewert, M. Pawlak, & M. Wrembel (Eds.), *EUROSLA Yearbook 13* (pp. 139-75). Amsterdam: John Benjamins.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*, 376-405.

Hofmeister, P., Jaeger, T.F., Arnon, I., Sag, I.A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgements. *Language and Cognitive Processes, 28*, 48-87.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120,* 901-931.

Hopp, H. (2013). The development of L2 morphology. *Second Language Research*, *29*, 3-6.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346-363.

Howard, J.H., Jr., Howard, D.V., Dennis, N.A. & Kelly, A.J. (2008). Implicit learning of predictive relationships in three element visual sequences by young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1139-1157.

Hulstijn, J.H. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H.W. Dechert & M. Raupach (Eds.), *Interlingual processing* (pp. 49-73). Tubingen: Gunter Narr.

Hulstijn, J.H. (1997). Second language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition*, *19*, 131-143.

Hulstijn, J.H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193-223.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C.J. Doughty & M.H. Long (Eds.), *The Handbook of second language acquisition* (pp. 349-381). Malden, MA: Blackwell.

Hulstijn, J.H. (2007). Psycholinguistic perspectives on second language acquisition. In J. Cummins & C. Davison (Eds.), *The international handbook on English language teaching, Volume II, Current Trends and Issues* (pp. 701-713). Norwell, MA: Springer.

Hulstijn, J.H. (2013). Incidental learning in second language acquisition. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2632-2636). Chichester, West Sussex, UK: Wiley-Blackwell.

Hulstijn, J.H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view.

In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-46). Amsterdam: John Benjamins.

Hulstijn, J.H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, *34*, 23-43.

Hulstijn, J.H., Young, R.F., Ortega, L., Bigelow, M., DeKeyser, R.M., Ellis, N.C., Lantolf, J.P., Mackey, A. & Talmy, S. (2014). Bridging the gap: Cognitive and social approaches to research in second language learning and teaching. *Studies in Second Language Acquisition*, *36*, 361-421.

Ionin, T., & Zyzik, E. (2014). Judgement and interpretation tasks in second language research. *Annual Review of Applied Linguistics*, *34*, 37-64.

Isemonger, I.M. (2007). Operational definitions of explicit and implicit knowledge: Response to R. Ellis (2005) and some recommendations for future research in this area. *Studies in Second Language Acquisition*, *29*, 101-118.

Jackson, C.N. (2007). The use and non-use of semantic information, word order, and case markings during comprehension by L2 learners of German. *The Modern Language Journal*, *91*, 418-431.

Jackson, C.N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence procession. *Language Learning*, *58*, 875-909.

Jackson, D.O. (2013). Implicit second language learning and individual differences. In J.M. Bergsleithner, S.N. Frota & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 271-287). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Jackson, D.O. (2014). *The relative contribution of input modification, learner awareness, and individual differences to second language construction learning.* (Unpublished doctoral dissertation). University of Hawai'i at Manoa, Manoa, HI.

Jacob, G., & Felser, C. (in press). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59, 434-446.

James, W. (1890). *The Principles of Psychology*. New York: Holt.

Janda, L.A., & Townsend, C.E. (2000). Stand alone Czech. Munich: Lincom.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, *50*, 245-309.

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

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20-49). New York: Routledge.

Jeon, H.E. (2015). Multiple regression. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 131-158). New York: Routledge.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*, 1-33.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, *21*, 207-220.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199-225.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*, 121-151.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*, 137-166.

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *98*, 122-149.

Just, M.A., Carpenter, P.A., & Wooley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228-238.

Kaiser, E. (2013). Experimental paradigms in psycholinguistics. In R.J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 135-168). Cambridge: Cambridge University Press.

Karuza, E., Farmer, T.A., Smith, F.X., Fine, A.B., & Jaeger, T.F. (2014). Relating changes in on-line processing time to learning outcomes in a self-paced statistical learning task. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 725-730). Austin, TX: Cognitive Science Society.

Kaufman, S.B., DeYoung, C.G., Gray, J.R., & Jimenéz, L., Brown, J.B., & Mackintosh, N. (2010). Implicit learning as an ability, *Cognition*, *116*, 321-340.

Kaufman, S.B., DeYoung, C.G., Reis, D.L., & Gray, J.R. (2011). General intelligence predicts reasoning ability even for evolutionarily familiar content. *Intelligence*, *39*, 311-322.

Keating, G.D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*, 1-32.

Keenan, J.M., & MacWhinney, B. (1987). Understanding the relationship between comprehension and production. In H.W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 149-155). Norwood, N.J.: Ablex Publishing Corporation.

Kempe, V., & Brooks, P.J. (2001). The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning*, *51*, 221-256.

Kempe, V. & Brooks, P.J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*, 703-746.

Kempe, V., Brooks, P.J., & Gillis, S. (2007). Diminutives provide multiple benefits for language acquisition. In W. Dressler & I. Savickiene (Eds.), *The acquisition of diminutives: A cross-linguistic perspective* (pp. 319-342). Amsterdam: John Benjamins.

Kempe, V., Brooks, P.J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*, 127-153.

Kempe, V., Brooks, P.J., & Pirott, L. (2001). How can child-directed speech facilitate the acquisition of morphology? In M. Almgren, A. Barrena, M.-J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (Eds.), *Research on Child Language Acquisition: Proceedings of the VIIIth International Congress for the Study of Child Language* (pp. 1237-1247). Medford, MA: Cascadilla Press.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, *20*, 543-587.

Kielar, A., Meltzer, J.A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory responses to semantic and syntactic violations. *Journal of Cognitive Neuroscience*, *26*, 2840-2862.

Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, *16*, 776-791.

Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, *13*, 301-347.

Kliegl, R., Masson, M.E., & Richter, E.M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*, 655-681.

Krashen, S.D. (1978). Individual variation in the use of the monitor. In W. Ritchie (Ed.), *Second language acquisition research* (pp. 175-183). New York: Academic Press.

Krashen, S.D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

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

Krashen, S.D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Krashen, S.D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Krashen, S.D. (2003). *Explorations in language acquisition and use*. Portsmouth: Heinemann.

Krause, H., Bosch, S., & Clahsen, H. (in press). Morphosyntax in the bilingual mental lexicon. *Studies in Second Language Acquisition*.

Lachaud, C.M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics, 32*, 389-416.

Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, *60*, 221-230.

Larsen-Freeman, D. (2013). Transfer of learning transformed. *Language Learning*, *63*, 107-129.

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, *65*, 127-159.

Lee, J.F. (2015). The milestones in twenty years of processing instruction research. *International Review of Applied Linguistics in Language Teaching*, *53*, 111-126.

Lee, J.F., & Benati, A. (2013). *Individual differences and processing instruction*. London: Equinox.

Leeser, M.J. (2014). On psycholinguistic methods. In J. Jegerski, & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 231-251). New York: Routledge.

Leow, R.P. (1997). Attention, awareness, and foreign language behaviour. *Language Learning, 47*, 467-505.

Leow, R.P. (2000). A study of the role of awareness in foreign language behaviour: Aware versus unaware learners. *Studies in Second Language Acquisition*, *22*, 557-584.

Leow, R.P. (2001). Attention, awareness, and foreign language behaviour. *Language Learning*, *51*, 113-155.

Leow, R.P. (2002). Models, attention, and awareness in SLA: A response to Simard and Wong's (2001) 'Alertness, orientation, and detection: The conceptualization of attentional functions'. *Studies in Second Language Acquisition, 24*, 113-119.

Leow, R.P. (2012). Explicit and implicit learning in the L2 classroom: What does the research suggest? *The European Journal of Applied Linguistics and TEFL*, *2*, 1-14.

Leow, R.P. (2013). Schmidt's noticing hypothesis: More than two decades after. In J. Bergsleithner, S. Frota, & J. Yoshioka (Eds.) *Noticing: L2 studies and essays in honor in Dick Schmidt* (pp. 11-24). Honolulu, HI: The National Foreign Language Resource Center.

Leow, R.P. (2015a). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 47-66). Amsterdam: John Benjamins.

Leow, R.P. (2015b). *Explicit learning in the L2 classroom: A student-centred approach*. New York: Routledge, Taylor, & Francis Group.

Leow, R.P., & Hama, M.K. (2013). Implicit learning in SLA and the issue of internal validity: A response to Leung and Williams' 'The implicit learning of mappings between forms and contextually derived meanings'. *Studies in Second Language Acquisition*, *35*, 545-557.

Leung, J.H.C., & Williams, J.N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, *33*, 33-55.

Leung, J.H.C., & Williams, J.N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*, 634-662.

Leung, J.H.C., & Williams, J.N. (2014). Crosslinguistic differences in implicit language learning. *Studies in Second Language Acquisition*, *36*, 733-755.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology, 20*, 24-37.

Linck, J.A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185-207.

Linck, J.A., Kroll, J.F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*, 1507-1515.

Lightbrown, P.M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z-H. Han (Ed.), *Understanding second language process* (pp. 27-44). Clevedon, UK: Multilingual Matters.

Loewen, S. (2009). Grammaticality judgement tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 31-64). Bristol, UK: Multilingual Matters.

Loewen, S. (2015). *Instructed second language acquisition*. New York: Routledge.

Long, M.H. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition* (pp. 114-141). New York: Newbury House.

Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Leeds: Emerald Group Publishing Limited.

Long, M.H. (2007). *Problems in SLA*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Long, M.H. (2013). Maturational constraints on child and adult SLA. In G. Granena & M.H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 3-41). Amsterdam: John Benjamins.

Long, M.H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.

Lukavský, J., & Smolík, F. (2009). Word order and case inflection in Czech: On-line sentence comprehension in children and adults. In *Proceedings of the 31st Conference of the Cognitive Science Society* (pp. 1358-1363).

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*, 405-530.

Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 21-46). Cambridge: Cambridge University Press.

Mackey, A. & Gass, S.M. (2005). *Second language research methodology and design*. Mahwah, N.J.: Lawrence Erlbaum Associates.

MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. *Journal of Psycholinguistic Research*, *12*, 467-478.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd Edition). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*, 144-161.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In S. Unsworth & E. Blom (Eds.), *Experimental methods in language acquisition research* (pp. 139-162). Amsterdam: John Benjamins.

Marsden, E.J. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning*, *56*, 507-566.

Marsden, E.J., Altmann, G., & St. Claire, M. (2013). Priming of verb inflections in L1 and L2 French: A comparison of 'redundant' versus 'non-redundant' training conditions. *International Review of Applied Linguistics in Language Teaching*, *51*, 271-298.

Marsden, E.J., & Chen, H.Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, *61*, 1058-1098.

Marsden, E.J., Williams, J.N., & Liu, X. (2013). Learning novel morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, *35*, 619-654.

Mathews, R.C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General*, *119*, 412-416.

McLaughlin, B. (1978). The monitor model: Some methodological considerations. *Language Learning*, *28*, 309-332.

McLaughlin, B. (1980). On the use of miniature artificial languages in second-language research. *Applied Psycholinguistics*, *1*, 357-369.

McLaughlin, B. (1990a). "Conscious" versus "unconscious" learning. *TESOL Quarterly*, *24*, 617-634.

McLaughlin, B. (1990b). Restructuring. *Applied Linguistics, 11*, 113-128.

Merikle, P.M. (1982). Unconscious perception revisited. *Perception & Psychonomics*, *31*, 298-301.

Merikle, P.M. (1992). Perception without awareness: Critical issues. *American Psychologist*, *47*, 792-795.

Misyak, J.B., & Christiansen, M.H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*, 302-331

Misyak, J.B., Christiansen, M.H., & Tomblin, J.B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*, 138-153.

Misyak, J.B., Goldstein, M.H., Christiansen, M.H. (2012). Statistical-sequential learning in development. In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 13-54). Berlin: Mouton de Gruyter.

Mitchell, D.C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D.E. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 69-89). Hillsdale, N.J..: Lawrence Erlbaum Associates.

Mitchell, D.C., & Green, D.W. (1978). The effects of context and content on immediate processing in reading. *The Quarterly Journal of Experimental Psychology*, *30*, 609-636.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. New York: Routledge.

Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A.F. Healy & L.E. Bourne (Eds.), *Foreign* *language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, N.J.: Lawrence Erlbaum Associates.

Moeser, S.D., & Bregman, A.S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behaviour, 11*, 759-769.

Monaghan, P., Christiansen, M.H., & Fitneva, S.A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*, 325-347.

Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, *23*, 564-579.

Morgan, J.L., & Newport, E.L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behaviour, 20*, 67-85.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*, 154-193.

Morgan-Short, K., & Tanner, D. (2014). Event-related potentials (ERPs). In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 127-152). New York: Routledge.

Muma, J.R. (1993). The need for replication. *Journal of Speech, Language, and Hearing Research*, *36*, 927-930.

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *IRAL-International Review of Applied Linguistics in Language Teaching*, *46*, 197-220.

Nagel, T. (1974).What is it like to be a bat? *The Philosophical Review*, *83*, 435-450.

Nassaji, H. (2012). Statistical significance tests and results generalizability: Issues, misconceptions, and a case for replication. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 92-115). Cambridge: Cambridge University Press.

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language. *Studies in Second Language Acquisition*, *31*, 403-435.

Neumann, O. (1996). Theories of attention. In O. Neumann & W. Prinz (Eds.), *Handbook of perception and action Vol. 3: Attention* (pp. 289-446). San Diego: Academic Press.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.

Norman, E., Price, M., Duff, S., & Mentzoni, R. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, *16*, 809-837.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417-528.

Ortega, L. (2009). Sequences and processes in language learning. In M.H. Long and C. Doughty (Eds.), *The handbook of language teaching* (pp. 81-105). Oxford: Wiley–Blackwell.

Osborne, J. (2005). Notes on the use of data transformations. *Practical Assessment, Research and Evaluation*, *9*, 42-50.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199-230.

Overgaard, M., Timmermans, B., Sandberg, K., & Cleeremans, A. (2010). Optimizing subjective measures of consciousness. *Consciousness and Cognition*, *19*, 682-684.

Paciorek, A., & Williams, J.N. (2015a). Semantic implicit learning. In P. Rebuschat (Ed.) *Implicit and explicit learning of languages* (pp. 67-88). Amsterdam: John Benjamins.

Paciorek, A., & Williams, J.N. (2015b). Implicit learning of semantic preferences of verbs. *Studies in Second Language Acquisition*, *37*, 359-382.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 80-96.

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Maidenhead, Berkshire, England: McGraw Hill.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (393-419). London: Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Park, E.S. (2014). Learners' first exposure to written L2 input: Insights from learners' spontaneous think-alouds. In Z-H. Han & R. Rast (Eds.), *First exposure to a second language: Learners' initial input processing* (pp. 40-63). Cambridge: Cambridge University Press.

Park, E.S., & Han, Z-H. (2008). Learning spontaneous attention in second language input processing: An exploratory study. In Z-H. Han (Ed.), *Understanding second language process* (pp. 106-132). Clevedon: Multilingual Matters.

Payne, J.W. (1994). Thinking aloud: Insights into information processing. *Psychological Science*, *5*, 241-248.

Penke, M. (2012). The Dual-mechanism debate. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 574-595). Oxford: Oxford University Press.

Perruchet, P. (2008). Implicit learning. In H.L. Roediger (Ed.), *Cognitive psychology of memory, Volume 2* (pp. 597-621). Oxford, UK: Elsevier.

Perruchet, P., Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. *Quarterly Journal of Experimental Psychology, 44*, 193-210.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology*, *22*, 493-516.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*, 264-275.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233-238.

Perry, F. (2011). *Research in applied linguistics: Becoming a discerning consumer*. New York: Routledge.

Peters, A. (1985). Language segmentation: Operating principles for the perception and analysis of language. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. II, pp. 1029-1067). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Petersson, K.M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*, 83-95.

Petr, Jan. (Ed.) (1986). *Mluvnice češtiny*. Prague: Academia Press.

Pinheiro, J.C., & Bates, D.M. (2000). *Mixed-effects models in S and S-PLUS (Statistics and Computing).* New York: Springer.

Pinker, S. (1999). How the mind works. *Annals of the New York Academy of Sciences*, *882*, 119-127.

Pinker, S., & Ullman, M.T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456-463.

Plonsky, L. (2012). Replication, meta-analysis, and generalisability. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 116-132). Cambridge: Cambridge University Press.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655-687.

Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and call for reform. *Modern Language Journal*, *98*, 450-470.

Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes: A “back-to-basics” approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23-45). New York: Routledge

Plonsky, L., & Oswald, F.L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878-912.

Polio, C. (2012). Replication in published applied linguistics research: A historical perspective. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 47-91). Cambridge: Cambridge University Press.

Polio, C., & Gass, S.M. (1997). Replication and reporting. *Studies in Second Language Acquisition*, *19*, 499-508.

Porte, G. (2012). Introduction. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 1-18). Cambridge: Cambridge University Press.

Posner, M.I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173-206). Washington, DC: American Psychological Association.

Posner, M.I. (1990). Hierarchical distributed networks in the neuropsychology of selective attention. In A. Caramazza (Ed.), *Cognitive neuropsychology and neurolinguistics: Advances in models of cognitive function and impairment* (pp. 187-210). New York: Plenum Press.

Posner, M.I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science*, *7*, 11-14.

Posner, M.I., & Carr, T.H. (1992). Lexical access and the brain: Anatomical constraints on cognitive models of word recognition. *American Journal of Psychology*, *105*, 1-26.

Pothos, E.M. (2007). Theories of artificial grammar learning. *Psychonomic Bulletin*, *133*, 227-244.

Purpura, J.E., Brown, J.D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. *Language Learning*, *65*, 37-75.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413-425.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

Rast, R. (2008). *Foreign language input: Initial processing*. Clevedon, Multilingual Matters.

Rast, R. (2010). The use of prior linguistic knowledge in the early stages of L3 acquisition. *International Review of Applied Linguistics in Language Teaching*, *48*, 159-183.

Rast, R., Watorek, M., Hilton, H., & Shoemaker, E. (2014) Initial processing and use of inflectional markers: Evidence from French adult learners of Polish. In Z-H. Han and R. Rast (Eds.), *First exposure to a second language: learners' initial input processing* (pp. 64-106). Cambridge: Cambridge University Press.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510-532.

Rausch, M., Müller, H.J., & Zehetleitner, M. (in press). Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Consciousness and Cognition.*

Reber, A.S. (1965). *Implicit learning of artificial grammars* (Unpublished MA thesis). Brown University: Providence, RI.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 6*, 855-63.

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. J*ournal of Experimental Psychology: General*, *118*, 219-235.

Reber, A.S. (1990). On the primacy of the implicit: Comment on Perruchet and Pacteau. *Journal of Experimental Psychology: General*, *119*, 340-342.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* London: Clarendon Press.

Reber, A.S. (2011). An epitaph for grammar: An abridged history. In C. Sanz & R.P. Leow (Eds.) *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA & bilingualism* (pp. 23-34). Washington, DC: Georgetown 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., & Allen, R. (2009). Explicit versus implicit learning. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *The Oxford companion to consciousness* (pp. 403-407). Oxford: Oxford University Press.

Reber, A.S., Allen, R., & Reber, P.J. (1999). Implicit versus explicit learning. In R.J. Sternberg (Ed.), *The nature of cognition* (pp. 475-513). Cambridge, MA: MIT Press.

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. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioural Neuroscience*, *113*, 235-242.

Rebuschat, P. (2008). *Implicit learning of natural language syntax* (Unpublished dissertation). University of Cambridge, Cambridge, UK.

Rebuschat, P. (2013), Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595-626.

Rebuschat, P. (Ed.). (2015a). *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.

Rebuschat, P. (2015b). Introduction: Implicit and explicit learning of languages. In P. Rebuschat (Ed.), Implicit and explicit learning of languages (pp. xiii-xxii). Amsterdam: John Benjamins.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K. & Ziegler, N. (2013) Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. In J. Bergsleithner, S. Frota, & J.K. Yoshioka (Eds.), *Noticing: L2 studies and essays in honor of Dick Schmidt* (pp. 255-275). Honolulu, HI: University of Hawai'i Press.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, *37*, 299-334.

Rebuschat, P., & Williams, J.N. (2006). Dissociating implicit and explicit learning of natural language syntax. In R. Sun. & N. Miyake (Eds.), *Proceedings of the annual meeting of the Cognitive Science Society* (p. 2594). Mahwah, N.J.: Lawrence Erlbaum Associates.

Rebuschat, P., & Williams, J.N. (2009). Implicit learning of word order. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 425-430). Austin, TX: Cognitive Science Society.

Rebuschat, P., & Williams, J.N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*, 829-856.

Reingold, E.M., & Merikle, P.M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44*, 563-575.

Reingold, E.M., & Merikle, P.M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language*, *5*, 9-28.

Reingold, E.M., & Merikle, P.M. (1993). Theory and measurement in the study of unconscious processes. In M. Davies & G. Humphreys (Eds.), *Consciousness* (pp. 40-57). Oxford: Basil Blackwell Ltd.

Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology*, *61*, 400-424.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*, 93-132.

Riefer, D.M., & Rouder, J.N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*, 601-611.

Roberts, L. (2012a). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*, 113-127.

Roberts, L. (2012b). Individual differences in second language sentence processing. *Language Learning*, *62*, 172-188.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*, 299-331.

Roberts, L., & Liszka, S.A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, *29*, 413-439.

Robertson, E.M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27*, 10073-10075.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*, 283-331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27-67.

Robinson, P. (1997). Generalisability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*, 223-247.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfeld and

Hernstadt, 1991. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*, 235-268.

Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: how comparable are they? *Language Learning*, *60*, 245-263.

Robinson, P., Mackey, A., Gass, S.M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S.M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*, (pp. 247-267). New York: Routledge.

Roehr, K. (2008). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics*, *19*, 67-106.

Roehr, K. (2010). *Explicit knowledge and learning in SLA*. Amsterdam: John Benjamins.

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, *49*, 857-866.

Rogers, J., Révész, A., & Rebuschat, P. (2015). Challenges in implicit learning research: Validating a novel artificial language. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 273-298). Amsterdam: John Benjamins.

Rosa, E.E., & Leow, R.P. (2004). Awareness, different learning conditions, and L2 development. *Applied Psycholinguistics*, *25*, 269-292.

Rosa, E.E., & O'Neill, M.D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition*, *21*, 511-556.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Sagarra, N. (2008). Working memory and L2 processing of redundant grammatical forms. In Z-H. Han (Ed.), *Understanding second language process* (pp. 133-147). Clevedon: Multilingual Matters.

Sagarra, N. (2013). Working memory in second language acquisition. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 6207-6215). Oxford, UK: Wiley–Blackwell.

Sagarra, N., & Ellis, N.C. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, *35*, 261-290.

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, *61*, 80-116.

Sandberg, K., Bibby, B.M., & Overgaard, M. (2013). Measuring and testing awareness of emotional face expressions. *Consciousness and Cognition*, *22*, 806-809.

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, *19*, 1069-1078.

Sanz, C., & Morgan-Short, K. (2005). Explicitness in pedagogical interventions: Input, practice, and feedback. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 234-263). Washington, DC: Georgetown University Press.

Sanz, C., & VanPatten, B. (1998). On input processing instruction, and the nature of replication tasks: A response to Salaberry. *Canadian Modern Language Review*, *54*, 263-273.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace, & World.

Schacter, D.L. (1989). On the relations between memory and consciousness: Dissociable interactions and conscious experience. In H.L. Roediger and F.I.M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 355-389). Mahwah, N.J.: Lawrence Erlbaum Associates.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129-58.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13,* 206-226

Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. In J. H. Hulstijn, & R. Schmidt (Eds.) *Consciousness and second language learning: Conceptual, methodological and practical issues in language learning and teaching, Thematic issue of AILA Review - Revue de l'AILA*, *11,* 11-26.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). London: Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: University of Hawai'i Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W.M. Chan, S. Chi, K.N. Cin, J. Instanto, M. Nagami, J.W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721-37). Singapore: National University of Singapore, Centre for Language Studies.

Schmidt, R.W., & Frota, S.N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R.R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schütze, C.T. (2001). On the nature of default case. *Syntax*, *4*, 205-238.

Schütze, C.T., & Sprouse, J. (2013). Judgement data. In R.J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27-50). Cambridge: Cambridge University Press.

Searle, J. (1983, October). The world turned upside down. *The New York Review of Books*, *30*(16), 74-79.

Searle, J. (2007). Biological naturalism. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness*. Malden, MA: Blackwell.

Seidenberg, M.S., & Elman, J.L. (1999). Networks are not ‘hidden rules’. *Trends in Cognitive Sciences*, *3*, 288-289.

Segal, E.M., & Halwes, T.G. (1965). Learning of letter pairs as a prototype of first language learning. *Psychonomic Science*, *3*, 451-452.

Segal, E.M., & Halwes, T.G. (1966). The influence of frequency of exposure on the learning of a phrase structural grammar. *Psychonomic Science*, *4*, 157-158.

Serafini, E. (2013). *Cognitive and psychological factors in the long-term development of implicit and explicit second language knowledge in adult learners of Spanish at increasing proficiency* (Unpublished dissertation). Georgetown University, Washington, DC.

Servan-Schreiber, E., & Anderson, J.R. (1990) Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 592-608.

Seth, A.K., Baars, B.J, & Edelman, D.B. (2005). Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, *14*, 119-139.

Siegler, R.S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*, 606-620.

Shanks, D.R. (2005). Implicit learning. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202-220). London: Sage.

Shanks, D.R., & St. John, M.F. (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences*, *17*, 367-447.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, *2*, 159-169.

Sharwood Smith, M. (2008). Revisiting the role of consciousness with MOGUL. In Z-H. Han (Ed). *Understanding Second Language* Processes (pp. 1-15) Clevedon: Multilingual Matters.

Sharwood Smith, M. (2015). A commentary on processing instruction. *International Review of Applied Linguistics in Language Teaching*, *53*, 271-275.

Shirai, Y. (2002). The prototype hypothesis of tense-aspect acquisition in second language. *Language Acquisition and Language Disorders*, *27*, 455-47.

Shirai, Y. (2004). A multiple-factor account for the form-meaning connections in the acquisition of tense-aspect morphology. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 91-112). Mahwah, N.J.: Lawrence Erlbaum Associates.

Short, D. (2009) Czech and Slovak. In B. Comrie (Ed.), *The world's major languages* (2nd Ed., pp. 305-329). New York: Routledge.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*, 245-260.

Slobin, D.I. (1997). The universal, the typological, and the particular in acquisition. In D.I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 5, pp.1-37). Mahwah: Lawrence Erlbaum.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.) *The Crosslinguistic Study of Language Acquisition, Vol. II* (pp. 1157-1256). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Smallwood, J., & Schooler, J.W. (2009). Mind-wandering. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *The Oxford companion to consciousness* (pp. 443-445). Oxford: Oxford University Press.

Smith, K.H. (1966). Grammatical intrusions in the recall of structured letter pairs: Mediated transfer or position learning? *Journal of Experimental Psychology*, *72*, 580-588.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121-159.

Song, J.J. (2011). Word order typology. In J.J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp.253-279). Oxford: Oxford University Press.

Squire, L.R., & Knowlton, B.J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 765-779). Cambridge, MA: MIT Press.

Stump, G.T. (2001). Inflection. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology* (pp. 13-43). Oxford: Blackwell.

Tabachnick, B.G., & Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

Tagarelli, K. M. (2014). *The neurocognition of adult second language acquisition: An fmri study* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Tagarelli, K.M., Borges Mota, M., & Rebuschat, P. (2011). The role of working memory in the implicit and explicit learning of languages. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2061-2066). Austin, TX: Cognitive Science Society.

Tagarelli, K.M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning context, and the acquisition of L2 syntax. In W. Zhisheng., M. Borges Mota., & A. McNeill (Eds.), *Working memory in second language acquisition and processing: Theory, research and commentary* (pp. 224-247). Bristol: Multilingual Matters.

Tagarelli, K.M., Jiang, X., Laka, I., Barbey, A.K., Morgan-Short, K., Ullman, M.T. (2014, March). Examining the trajectory of language acquisition with a mini-language model.

Paper presented at the annual meeting of the AAAL Annual Conference. Portland, Oregon.

Timmermans, B. & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 21-46). Oxford: Oxford University Press.

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Linguistics*, *22*, 1092-1106.

Tomlin, R.S., & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition*, *16*, 183-203.

Traxler, M.J., & Tooley, K.M. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes*, *23*, 609-645.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research*, *14*, 103-135.

Truscott, J. (2015a). *Consciousness and second language learning*. Bristol: Multilingual Matters.

Truscott, J. (2015b). Consciousness in SLA: A modular perspective. *Second Language Research*, *31*, 413-434.

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition*, *33*, 497-528.

Ullman, M.T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105-122.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231-270.

Ullman, M.T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141-178). Washington, DC: Georgetown University Press.

Ullman, M.T. (2013). Declarative/procedural model of language. In H. Pashler (Ed.), *Encyclopedia of the mind* (pp. 224-226). Los Angeles: Sage Publications.

Vainio, S., Pajunen, A., & Hyönä, J. (2014). L1 and L2 word recognition in Finnish. *Studies in Second Language Acquisition*, *36*, 133-162.

Valdman, A. (1993). Replication study (editorial introduction). *Studies in Second Language Acquisition*, *15*, 505.

van den Bos, E., Christiansen, M.H., & Misyak, J.B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*, 507-520.

Van den Branden, K. (2007). Practice in perfect learning conditions? In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 161-179). Cambridge: Cambridge University Press.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, N.J.: Ablex.

VanPatten, B. (2002a). Processing instruction: An update. *Language Learning*, *52*, 755-803.

VanPatten, B. (2002b). Processing the content of input-processing and processing instruction research: A response to DeKeyser, Salaberry, Robinson, and Harrington. *Language Learning*, *52*, 825-831.

VanPatten, B. (Ed.) (2004a). Processing instruction: Theory, research, and commentary. Mahwah, N.J.: Lawrence Erlbaum Associates.

VanPatten, B. (2004b). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-32). Mahwah, N.J.: Lawrence Erlbaum Associates.

VanPatten, B. (2005). Processing instruction. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 267-281). Washington, DC: Georgetown University Press.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 115-136). Mahwah, N.J.: Lawrence Erlbaum Associates.

VanPatten, B. (2014). The psycholinguistics of SLA. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp.-19). New York: Routledge.

VanPatten, B. (2015a). Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching*, *53*, 91-109.

VanPatten, B. (2015b). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition: An introduction (2nd ed., pp. 113-134). New York: Routledge.

VanPatten, B., & Benati, A. G. (2010). *Key terms in second language acquisition*. London: Bloomsbury Publishing.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225-243.

VanPatten, B., & Williams, J. (Eds.). (2007). *Theories in second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

VanPatten, B., & Williams, J. (Eds.). (2015). *Theories in second language acquisition* (2nd Ed.). New York: Routledge.

Van Voorhis, C.R.W., & Morgan, B.L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*, 43-50.

Velmans, M. (1996).What and where are conscious experiences? In M. Velmans (Ed.) *The science of consciousness: Psychological, Neuropsychological, and clinical reviews* (pp. 63-95). London: Routledge.

Velmans, M. (1999). When perception becomes conscious. *British Journal of Psychology*, *4*, 543-566.

Velmans, M. (2009a). How to define consciousness—and how not to define consciousness. *Journal of Consciousness Studies*, *16*, 139-156.

Velmans, M. (2009b). *Understanding Consciousness*. New York: Routledge.

Velmans, M. & Schneider, S. (Eds.). (2007). *The Blackwell companion to consciousness.* Malden: Blackwell Publishers.

Velupillai, V. (2012). *An introduction to linguistic typology*. Amsterdam: John Benjamins.

Wells, G. (1985). *Language development in the pre-school years*. Cambridge: Cambridge University Press.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics*, *22*, 1-22.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475-482.

White, L. (2003). Fossilisation in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, *6*, 129-141.

White, L. (2009). Grammatical theory: Interfaces and L2 knowledge. In W. Ritchie & T. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 49-68). Leeds: Emerald Group Publishing Limited.

Whittlesea, B.W., & 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.

Wickens, C.D. (1989). Attention and skilled performance. In D.H. Holding (Ed.), *Human Skills* (2nd ed. pp. 71-105). New York: John Wiley.

Wierzchoń, M., Asanowicz, D., Paulewicz, B., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and Cognition*, *21*, 1141-1153.

Williams, J.N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, N.J.: Lawrence Erlbaum Associates.

Williams, J.N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269-304.

Williams, J.N. (2009). Implicit learning in second language acquisition. In W. Ritchie & T.K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-344). Leeds: Emerald Group Publishing Limited.

Williams, J.N. (2010). Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning*, *60*, 221-244.

Williams, J.N. (2012).Working memory and SLA. In S.M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427-441). Oxon: Routledge.

Williams, J.N. (2013). Attention, awareness, and noticing in language processing and learning. In J.M. Bergsleithner, S.N. Frota, & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 51-69). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

Williams, J.N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua, 118*, 522-553.

Williams, J.N., & Rebuschat, P. (2013). Statistical learning and syntax: What can be learned, and what difference does meaning make. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 237-264). Berlin: Mouton De Gruyter.

Winter, B., & Reber, A.S. (1994). Implicit learning and the acquisition of natural languages. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 115-145). London: Academic Press.

Wong, W. (2004). The nature of processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 33-63). Mahwah, N.J.: Lawrence Erlbaum Associates.

Zajonc, R.B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27.

Zelazo, P., Moscovitch, M., & Thompson, E. (Eds.). (2007). *The Cambridge handbook of consciousness.* Cambridge: Cambridge University Press.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37*, 457-486.

Ziori, E., & Dienes, Z. (2012). The time course of implicit and explicit concept learning. *Consciousness and Cognition*, *21*, 204-216.

Ziori, E., & Pothos, E.M. (2015). Artificial grammar learning: An introduction to key issues and debates. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 245-272). Amsterdam: John Benjamins.

APPENDICES

List of Appendices

Appendix A: Information Sheet and Consent Form Used in Pilot Studies 1 and 2, and Experiments 1 and 2

LANCASTER UNIVERSITY
Department of Linguistics and English Language

INFORMATION SHEET

As part of my Doctoral studies in the Department of Linguistics and English Language, I am carrying out a study investigating how native speakers of English learn foreign language vocabulary. The study will take place at Lancaster University in late January 2013. The information from this study will be used as part of my PhD thesis as well as for future publications.

I have approached you because you are a native speaker of English studying at Lancaster University.

For this study, you will listen to a number of English sentences that contain a single foreign vocabulary word. Based on the sentence, you will then have to choose a picture that matches the meaning of the foreign word.

If you have any queries about the study, please feel free to contact myself or my course supervisor, Andrea Revesz, who can be contacted on andrea.revesz@exchange.lancs.ac.uk or by phone on 01524 5 93212. You may also contact the Head of Department, Prof. Elena Semino, on 01524 594176.

Signed

Robert Johnson Rogers

r.rogers1@lancaster.ac.uk

Lancaster University
Lancaster LA1 4YL
United Kingdom
Tel: +44 (0)1524 593045
Fax: +44 (0)1524 843085
http://www.ling.lancs.ac.uk

UNIVERSITY OF LANCASTER

Department of Linguistics and English Language

Consent Form

Project title: Learning foreign language vocabulary under auditory conditions

1. I have read and had explained to me by **Robert Johnson Rogers** the Information Sheet relating to this project.

2. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

3. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.

4. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix B: Information Sheet and Consent Form Used in Experiment 3

Faculty of Children and Learning
Dean of Faculty Professor Richard Andrews
Head of Faculty Administration Mr Manjit Benning

Leading education and social research
Institute of Education
University of London

20 Bedford Way
London WC1H 0AL
Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
www.ioe.ac.uk

Sentence Comprehension in a Foreign Language:
A research project
January 31st to December 31st, 2014

Information for Participants.

Please will you help with my research?

My name is John Rogers. I am currently a PhD student at the Institute of Education.

This leaflet tells you about my research.

I hope the leaflet will also be useful, and I
would be pleased to answer any questions you have.

Why is this research being done?

My research is focused on the general area of how we learn foreign languages. Specifically, this research investigates different factors which affect our ability to read and understand sentences in a foreign language.

Faculty of Children and Learning
Dean of Faculty Professor Richard Andrews
Head of Faculty Administration Mr Manjit Benning

Leading education and social research
Institute of Education
University of London

20 Bedford Way
London WC1H 0AL
Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
www.ioe.ac.uk

Who will be in the project?

This experiment is designed for native speakers of English. In particular, I have sought volunteers from different colleges in London.

What will happen during the research?

This entire experiment is computer based. You will complete a number of experiments on the computer, which are designed to investigate how native speakers of English read and understand sentences in a foreign language.

At the end of the experiment, I will briefly interview you about your experience during this experiment.

This experiment will take about 2 hours in total to complete (including a short break).

What will happen to you if you take part?

In the first part of the experiment, you will read a number of sentences and answer comprehension questions based on the sentences.

The second stage of the experiment consists of a number of shorter experiments which are designed to measure things like your short-term memory span and attention.

Faculty of Children and Learning
Dean of Faculty Professor Richard Andrews
Head of Faculty Administration Mr Manjit Benning

Leading education and social research
Institute of Education
University of London

20 Bedford Way
London WC1H OAL
Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
www.ioe.ac.uk

Could there be problems for you if you take part?

All details related to your involvement in this experiment will be anonymous. I will assign you a random participant number at the beginning of the experiment, which will be the only identifying feature.

If you have any problems with the project, please tell me.

Will doing the research help you?

I hope you will enjoy helping me in my research.

The research will help me learn to be a researcher so that I may do more research in the future which will help other people.

I am offering a £10 reward for your time and participation.

Who will know that you have been in the research?

I will keep notes in a safe place, and will change all the names in my reports – and the name of the setting– so that no one knows who said what.

Faculty of Children and Learning
Dean of Faculty Professor Richard Andrews
Head of Faculty Administration Mr Manjit Benning

Leading education and social research
Institute of Education
University of London

20 Bedford Way
London WC1H OAL
Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
www.ioe.ac.uk

Page 4

Do you have to take part?

Participation in this project is completely voluntary.

You decide if you want to take part and, even if you say 'yes', you can drop out at any time or say that you don't want to answer some questions.

You can tell me that you will take part by signing the consent form.

The project has been reviewed by the Faculty Research Ethics Committee.

Thank you for reading this leaflet.

John Rogers
rrogers@ioe.ac.uk

Faculty of Children and Learning
Dean of Faculty Professor Richard Andrews
Head of Faculty Administration Mr Manjit Benning

IOE LONDON

Leading education and social research
Institute of Education
University of London

20 Bedford Way
London WC1H OAL
Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
www.ioe.ac.uk

Consent form

Sentence Comprehension in a Second Language
January 31st to December 31st, 2014

I have read the information leaflet about the research. ☐ (please tick)

I agree to take part in this experiment. ☐ (please tick)

Name ________________________

Signed ________________________ date ______________

Researcher's name Robert Johnson Rogers

Signed ________________________ date ______________

Appendix C: Training Sets for Pilot Studies 1 and 2, and Experiments 1 and 2
Training Set

Nominative -a

PS1	PS2	E1	E2	
☑	☑	X	X	Today the way showed the smerovka on the street.
☑	☑	X	X	Last week the brasna broke its handle at the shop.
☑	☑	X	X	All night the boy frightened the zruda in his dreams.
☑	☑	☑	☑	In summer the pilka cut the wood with its teeth.
☑	☑	X	X	Last weekend holes made the vrtacka with the bit.
☑	☑	☑	☑	Last year the bedna stored pictures in the garage.
☑	☑	☑	☑	Today David's face cut the britva at the sink.
☑	☑	☑	☑	Last weekend the kladka raised the box to the roof.
☑	☑	X	X	Last night his mouth burned the bylina in the food.
☑	☑	☑	☑	All day the mycka did the washing with water.
☑	☑	☑	☑	Last week the door opened the klicka to the room.
☑	☑	X	X	All night the lednicka cooled the food at the store.
☑	☑	☑	☑	Last month him entertained the loutka with a show.
☑	☑	X	X	All year the prehrada supplied power to the town.
☑	☑	X	X	In the morning bread made the pekarna at home.
☑	☑	X	X	All year the zahrada grew veggies in the ground.
☑	☑	☑	☑	This morning the eggs moved the sterka to the plate.
☑	☑	☑	☑	Last year the kasa held Sarah's money at home.
☑	☑	☑	☑	Some time ago dirt cleaned the stetka in the sink.
☑	☑	☑	☑	Last weekend the kocka killed the bird with its teeth.
☑	☑	X	X	Last night the clothes pressed the zehlicka at his home.
☑	☑	☑	☑	This morning malta connected the stones in the wall.
☑	☑	☑	☑	Last winter them attacked the flota with its guns.
☑	☑	X	X	This evening the pracka pumped water to the clothes.
☑	☑	☑	☑	Last summer the grass ate the koza in the field.
☑	☑	X	X	All day the zkouska challenged them with hard questions.
☑	☑	☑	☑	Weeks ago the cheese took the myska to its hole.
☑	☑	X	X	All day the susicka dried clothes at the cleaners.
☑	☑	☑	☑	This morning them killed the flinta with a live round.
☑	☑	X	X	This morning the salina carried us to work.
☑	☑	☑	☑	At night a fire started a sirka at the camp.
☑	☑	X	X	This evening the guma made marks on the paper.
☑	☑	X	X	All summer phones built the tovarna with machines.
☑	☑	X	X	Last month the opona hid the stage at the show.
☑	☑	X	X	Last month a hole dug the lopata in the ground.
☑	☑	☑	☑	Last week the rtenka added colour to her face.
☑	☑	X	X	All evening the bath filled the sprchka in the home.
☑	☑	☑	☑	All week the lecba helped her in the hospital.
☑	☑	X	X	At night light made the svitilna in the garden.
☑	☑	☑	☑	Last winter the studna supplied water to town.

☑	☑	☑	X	In the morning the ship kept the kotva in place.
☑	☑	☑	X	This evening the krava fed its young with its milk.
☑	☑	X	X	All week lunch supplied the katedra to the kids.
☑	☑	X	X	Some time ago the salupa raised sails at sea.
☑	☑	☑	X	Last weekend the meat cut the dyka with its edge.
☑	☑	☑	X	Last year the delba carried water to his home.
☑	☑	X	X	All night wine opened the vyvrtka at the bar.
☑	☑	X	X	Last year the ohrada kept the cows in the field.

Category 2: Accusative -u

PS1	PS2	E1	E2	
☑	☑	X	X	In summer they placed the smerovku at the road.
☑	☑	X	X	Last weekend the brasnu put David in the boot.
☑	☑	X	X	Last week he drew a scary zrudu with the paint.
☑	☑	☑	☑	Last month the pilku pressed the builder to the board.
☑	☑	X	X	All week the builder took his vrtacku to work.
☑	☑	☑	☑	Some time ago the bednu closed Brian with tape.
☑	☑	☑	☑	Today Peter used a britvu in the bathroom.
☑	☑	☑	☑	All day a kladku used he at the building site.
☑	☑	X	X	At night Lucy cooked the bylinu with some oil.
☑	☑	☑	☑	Today the mycku connected James to the pipe.
☑	☑	☑	☑	In the morning she turned the klicku at the gate.
☑	☑	X	X	Last week the lednicku closed Jill in the kitchen.
☑	☑	☑	☑	Last night Katie carried the loutku to the stage.
☑	☑	X	X	Today the prehradu closed they at the river.
☑	☑	X	X	All year the baker used a pekarnu at work.
☑	☑	X	X	Weeks ago the zahradu planted Beth with fruits.
☑	☑	☑	☑	All week the cook used a sterku at the cooker.
☑	☑	☑	☑	Last month the kasu opened David with the key.
☑	☑	☑	☑	This morning the cleaner took the steku to work.
☑	☑	☑	☑	All morning the kocku rubbed she in the garden.
☑	☑	X	X	Today she pressed the zehlicku against her shirts.
☑	☑	☑	☑	Last weekend the maltu mixed Brian with a stick.
☑	☑	☑	☑	At night they positioned the flotu at the front.
☑	☑	X	X	Last week the pracku carried we to the kitchen.
☑	☑	☑	☑	Some time ago John kept the kozu at the farm.
☑	☑	X	X	Last week a zkousku gave the teacher to the class.
☑	☑	☑	☑	In summer the cat chased the mysku in the house.
☑	☑	X	X	Last year the susicku repaired Mark with his tools.
☑	☑	☑	☑	In the morning James fired the flintu in the air.
☑	☑	X	X	Last month the salinu stopped Peter with the bell.
☑	☑	☑	☑	Last winter Sarah moved the sirku to the coal.
☑	☑	X	X	Some time ago the gumu left Jill on her desk.
☑	☑	X	X	Last year the owners closed the tovarnu in town.

☑	☑	X	X	Last night the oponu closed Anne in the bedroom.
☑	☑	X	X	This morning Beth used a lopatu in the dirt.
☑	☑	☑	☑	Today her rtenku applied she with a brush.
☑	☑	X	X	Weeks ago he fixed the sprchku in the bathroom.
☑	☑	☑	☑	All evening the lecbu gave the nurse with a shot.
☑	☑	X	X	Last night she held the svitilku to the darkness.
☑	☑	☑	☑	Some time ago a studnu dug she at their home.
☑	☑	☑	X	Last night the sailor raised the kotvu with a chain.
☑	☑	☑	X	In winter the kravu carried he to the farm.
☑	☑	X	X	All day she attended the katedru in her town.
☑	☑	X	X	Today a salupu rowed Jill in the river.
☑	☑	☑	X	This morning John sharpened the dyku with a stone.
☑	☑	☑	X	Some time ago the delbu fixed he with his tools.
☑	☑	X	X	Last week Anne handled a vyvrtku in the pub.
☑	☑	X	X	Last weekend an ohradu built they at the farm.

Category 3: Instrumental -ou

PS1	PS2	E1	E2	
☑	☑	X	X	Last week the drivers warned they with a smerovkou.
☑	☑	X	X	Today David carried the food with a brasnou.
☑	☑	X	X	Some time ago the child scared John with a zrudou.
☑	☑	X	☑	Today he cut the wooden board with a pilkou.
☑	☑	X	X	This morning some holes made James with a vrtacku.
☑	☑	X	☑	In the morning Jill packed her things with a bednou.
☑	☑	X	☑	In the morning her leg cut Anne with a britvou.
☑	☑	X	☑	Last month Brian lifted the rocks with a kladkou.
☑	☑	X	X	Last year the food salted Beth with the bylinou.
☑	☑	X	☑	All night Peter washed the dishes with a myckou.
☑	☑	X	☑	This morning the door opened she with a klickou.
☑	☑	X	X	Last week Sarah froze the food with a lednickou.
☑	☑	X	☑	All night the child entertained he with a loutkou.
☑	☑	X	X	Last month they blocked the river with a prehradou.
☑	☑	X	X	This morning some bread made we with our pekarnou.
☑	☑	X	X	Last year she grew enough food with her zahradou.
☑	☑	X	☑	This evening the meat turned David with a sterkou.
☑	☑	X	☑	Last weekend John hid his money with a kasou.
☑	☑	X	☑	All evening the toilet cleaned Mark with a stetkou.
☑	☑	X	☑	Some time ago he scared the birds with his kockou.
☑	☑	X	X	All evening her clothes pressed Jill with a zehlickou.
☑	☑	X	☑	Some time ago Brian built the wall with maltou.
☑	☑	X	☑	Last night the river guarded they with a flotou.
☑	☑	X	X	In the evening Anne washed her dress with a prackou.
☑	☑	X	☑	In winter the grass maintained Beth with a kozou.

☑	☑	X	X	Weeks ago he failed the students with a zkouskou.
☑	☑	X	☑	This evening the woman scared we with a myskou.
☑	☑	X	X	All night Lucy dried her clothes with a susickou.
☑	☑	X	☑	Some time ago a bird shot Sarah with a flintou.
☑	☑	X	X	All day Katie crossed the city with a salinou.
☑	☑	X	☑	Last night a fire started David with a sirkou.
☑	☑	X	X	Last weekend John fixed the mistake with a gumou.
☑	☑	X	X	Last year cars produced the town with a tovarnou.
☑	☑	X	X	Today he darkened the room with the oponou.
☑	☑	X	X	All evening a hole dug James with the lopatou.
☑	☑	X	☑	Last weekend she darkened her eyes with rtenkou.
☑	☑	X	X	All evening himself cleaned he with a sprchou.
☑	☑	X	☑	Weeks ago he improved his health with the lecbou.
☑	☑	X	X	At night the bedroom lit Jill with a svitikou.
☑	☑	X	☑	All day she supplied fresh water with a studnou.
☑	☑	X	X	This morning the ship secured Anne with a kotvou.
☑	☑	X	X	In summer Beth produced some milk with a kravou.
☑	☑	X	X	Last month the trip arranged students with the katedrou.
☑	☑	X	X	All week we sailed the river with a salupou.
☑	☑	X	X	Last winter the man murdered she with a dykou.
☑	☑	X	X	In summer they supplied water with a delbou.
☑	☑	X	X	All year the wine opened he with a vyvrtkou.
☑	☑	X	X	All year they guarded their land with an ohradou.

Appendix D: Testing Set Used in Pilot Studies 1 and 2, and Experiment 2

<u>Grammatical Items</u>

Last week the fruit washed the prodejna with water.
Some time ago the brana blocked the road to town.
Last weekend food burned the trouba in the kitchen.
All evening the hracka entertained them at home.
This evening the room cleaned the sluzka with water.
Today the dilna delivered food to their home.
All year wood chopped the sekera in the forest.
In the morning the bunda warmed him at the park.
All year a paku built Peter with a board.
This morning the man changed the plenu in the bathroom.
Weeks ago the stuhu tied the man to the gift.
All evening Mark bought an andulku at the shop.
In the morning the kasnu built Katie with stone.
This evening he wore a plastenku in the rain.
Last night the propisku pressed she on the paper.
This morning she changed the ropu at the garage.
All night pages printed they with a tiskarnou.
Today papers joined he with a sesivackou.
Last weekend the food mixed James with a vareckou.
This morning Anne wrote a letter with the kridou.
All evening the present tied he with a strunou.
All week we cleaned the tables with a prachovkou.
Today him examined the doctor with a lupou.
This morning they turned the ship with the kosatkou.

Ungrammatical Items

At night her bag took the cizinku to the train.
In summer the lisku ate the eggs at the farm.
At night him welcomed the letusku with a drink.
Last winter the chuvu cleaned the mess in the room.
All week the car lent the pujchovnou to the man.
In summer the zenou bought the dress at the shop.
Last night the feed ate the veverkou in the tree.
In the morning the babickou surprised them with cake.
Last winter the lichotka gave he to his wife.
This evening she reduced the cena at the shop.
All evening the smetana mixed they in the bowl.
In summer David cooked the kachna with the fire.
This morning a babovkou gave we to our friend.
Last year Beth grew a paprikou in the garden.
Last month a chalupou built she at the river.
This morning he heated the chatou with a fire.
Last weekend the box closed Lucy with the paska.
Some time ago they cleared the ice with the skrabka.
Last week his wife surprised he with a kvetina.
Last week we sharpened the knife with the ocilka.
This evening a shirt made David with bavlnu.
This morning we fixed the sound with a ladickou.
Weeks ago a path cut they with a macetu.
Last year the town produced milk with a mlekarnu.

Appendix E: Testing Set Used in Experiment 1

Grammatical Items

Last weekend the fruit washed the prodna with water.
Some time ago the brana blocked the road to town.
Last weekend food burned the trouba in the kitchen.
All evening the hracka entertained them at home.
This evening the room cleaned the sluzka with water.
Today the dilna delivered food to their home.
All weekend wood chopped the sekra in the forest.
In the morning the bunda warmed him at the park.
All year a paku built Peter with a board.
This morning the man changed the plenu in the bathroom.
Weeks ago the stuhu tied the man to the gift.
Some time ago Mark bought an andku at the shop.
In the morning the kasnu built Katie with stone.
This evening Patrick wore a plastku in the rain.
Last evening the propku pressed she on the paper.
This morning she changed the ropu at the garage.
All night the pages printed the tiskna at work.
Today papers joined the seska in the office.
All day vegetables mixed the varka in the bowl.
Today the krida drew a picture on the page.
All evening a strunu tied he to the present.
This evening the pracka pumped water to the clothes.
Today the lupu used a nurse at the clinic.
Today the captain turned the kostku on the ship.

Ungrammatical Items

At night her bag took the cizinku to the train.
In summer the lisku ate the eggs at the farm.
At night him welcomed the letusku with a drink.
Last winter the chuvu cleaned the mess in the room.
All week the car lent the punchnu to the man.
In summer the zenu bought the dress at the shop.
Last night the feed ate the veverku in the tree.
In the morning the babku surprised them with cake.
Last winter the lichtka gave he to his wife.
This evening she reduced the cena at the shop.
All evening the smetana mixed they in the bowl.
In summer David cooked the kachna with the fire.
This morning a babka gave we to our friend.
Last year Beth grew a papka in the garden.
Last month a chalpa built she at the river.
This morning he heated the chata with a fire.
Last month the box sealed the pasku in the bedroom.
Last night the skabku cleared the ice on the window.
Some time ago a kvetna gave he to his wife.
Last year the cilku sharpened the knife on the stone.
This evening she cleaned the bavna with some water.
Last weekend a ladka carried they to the room.
Last night a path cut the macku in the forest.
Some time ago they closed the mlecna in the town.

Appendix F: Stimuli for M/C Task in Experiment 3

<u>Nominative</u>

Today the way showed the smerka to the drivers.
Last week the brasna broke its handle at the shop.
All night the boy frightened the zruda in his dreams.
Some time ago holes made the vrtka in the wall.
Last evening his mouth burned the bakna in the food.
All evening the ledka cooled the food at the store.
All year the predka supplied power to the town.
In the morning bread made the pekna in the town.
All summer the hradka grew veggies in the ground.
Last night the clothes pressed the licka at the cleaners.
This evening the pracka pumped water to the clothes.
All day the kouska tested them with hard questions.
All weekend the suska dried their clothes in the sun.
This morning the lina carried us to the town.
Last night the guma erased marks on the pages.
This evening the pona hid the stage at the show.

<u>Accusative</u>

This morning Patrick used a lopku in the dirt.
Weeks ago he fixed the sprchku in the bathroom.
Last evening she held the svitku to the darkness.
All morning she attended the katdru in her town.
All evening a salpu rowed they in the river.
Last week Anne used a vrtku in the restaurant.
Some time ago a hradu built they at the farm.
In the morning he drove the holku to her school.
In the evening she took the vacku to the show.
Last week the police took the bandu to prison.
In the morning he hired a ridku in the street.
This morning he ordered the pajku to the front.
Some time ago James left the gumu on his desk.
In summer Anne put the tonku into the lock.
Last weekend Mark used the bartku in the garden.
Last weekend the child called her mamku on the phone.

Appendix G: Stimuli for Experiment 3

SPR Task

Training Sentences and Comprehension Questions

Sentence	In summer the pilka cut the wood with its teeth.
CQ1	Was the wood cut in winter?
CQ2	Was the plastic cut?
CQ3	Did it cut the wood with its teeth?
Sentence	Last year the bedna stored pictures in the garage.
CQ1	Were the pictures stored in the closet?
CQ2	Did this happen last year?
CQ3	Were the pictures kept on the wall?
Sentence	Today David's face cut the britva at the sink.
CQ1	Was David’s face injured?
CQ2	Did this happen last month?
CQ3	Was David's face cut in the bedroom?
Sentence	Last weekend the kladka raised the box to the roof.
CQ1	Was the box moved to the ground?
CQ2	Was the box lifted?
CQ3	Did this happen yesterday?
Sentence	Today the mycku connected James to the pipe.
CQ1	Did this happen today?
CQ2	Did John connect something to the pipes?
CQ3	Was something connected to the wall?
Sentence	In the morning she turned the klicku at the gate.
CQ1	Did she turn something at the door?
CQ2	Did she turn something?
CQ3	Did this happen in the evening?
Sentence	Last night Katie carried the loutku to the stage.
CQ1	Did she do this last night?
CQ2	Did something carry Katie?
CQ3	Was something carried to the stage?
Sentence	All week the cook used a sterku at the cooker.
CQ1	Did an actor use something?
CQ2	Was something used in the kitchen?
CQ3	Was something used at the cooker this morning?
Sentence	Last year the kasa held Sarah's money at home.
CQ1	Did this happen last year?
CQ2	Did something hold Sarah's car?
CQ3	Was Sarah's money held in a bank?
Sentence	Some time ago dirt cleaned the stetka in the sink.
CQ1	Was the bath cleaned by something?
CQ2	Was the dirt cleaned in the sink?
CQ3	Did this happen recently?

Sentence	Last weekend the kocka killed the bird with its teeth.
CQ1	Was a mouse killed?
CQ2	Did something kill a bird with its claws?
CQ3	Did this happen last weekend?
Sentence	This morning malta connected the stones in the wall.
CQ1	Were the stones connected by something?
CQ2	Did this happen a long time ago?
CQ3	Were the stones connected in the wall?
Sentence	At night they positioned the flotu at the front.
CQ1	Did this happen in the morning?
CQ2	Did something move them?
CQ3	Was something moved to the front?
Sentence	Some time ago John kept the kozu at the farm.
CQ1	Was something kept in the house?
CQ2	Did John keep something at the farm?
CQ3	Did this happen recently?
Sentence	In summer the cat chased the mysku in the house.
CQ1	Was something chased by the cat?
CQ2	Did this happen in winter?
CQ3	Did the cat chase something in the garden?
Sentence	In the morning James fired the flintu in the air.
CQ1	Did James fire something?
CQ2	Was something fired at a target?
CQ3	Did this happen in the afternoon?
Sentence	At night a fire started a sirka at the camp.
CQ1	Did a fire start something?
CQ2	Did this happen during the day?
CQ3	Was there a fire at the camp?
Sentence	Last week the rtenka added colour to her face.
CQ1	Was colour added to her clothes?
CQ2	Did something add colour to her face?
CQ3	Did this happen today?
Sentence	All week the lecba helped her in the hospital.
CQ1	Was she helped by something?
CQ2	Did something help her at home?
CQ3	Did something help her only on Monday?
Sentence	Last winter the studna supplied water to town.
CQ1	Did this happen in summer?
CQ2	Was water supplied to town?
CQ3	Did something supply electricity to town?
Sentence	Last night the sailor raised the kotvu with a chain.
CQ1	Did this happen last week?
CQ2	Did the soldier raise something?

CQ3	Was something raised by a chain?
Sentence	In winter the kravu carried he to the farm.
CQ1	Was he carried to the farm?
CQ2	Was something carried to the farm in winter?
CQ3	Did he carry something in summer?
Sentence	This morning John sharpened the dyku with a stone.
CQ1	Did John throw something?
CQ2	Was something sharpened?
CQ3	Did he sharpen something this evening?
Sentence	Some time ago the delbu fixed he with his tools.
CQ1	Did a woman fix something?
CQ2	Was something fixed with his tools?
CQ3	Was something fixed recently?
Sentence	All day the mycka did the washing with water.
CQ1	Did something do the washing?
CQ2	Was the washing done only in the morning?
CQ3	Was the washing done with water?
Sentence	Last week the door opened the klicka to the room.
CQ1	Was the door to the basement opened?
CQ2	Did something open the door?
CQ3	Did something open the door last month?
Sentence	Last month him entertained the loutka with a show.
CQ1	Was he entertained by something?
CQ2	Was he entertained with a book?
CQ3	Did this happen this morning?
Sentence	This morning the eggs moved the sterka to the plate.
CQ1	Were the eggs moved this afternoon?
CQ2	Did something move the eggs to the plate?
CQ3	Was something moved to the pan?
Sentence	Last month the pilku pressed the builder to the board.
CQ1	Did he press something against the board?
CQ2	Was something pressed against the board this morning?
CQ3	Was something pressed against the board?
Sentence	Some time ago the bednu closed Brian with tape.
CQ1	Did Brian do something?
CQ2	Did he use tape for something?
CQ3	Did this happen recently?
Sentence	Today Peter used a britvu in the bathroom.
CQ1	Did Peter use something in the bathroom?
CQ2	Was something used in the kitchen?
CQ3	Was something used in the bathroom by Peter today?
Sentence	All day a kladku used he at the building site.
CQ1	Was something used in the office all day?

CQ2	Did he use something at the building site?
CQ3	Did he use something all day?
Sentence	Last winter them attacked the flota with its guns.
CQ1	Were they attacked by something?
CQ2	Did something attack them with arrows?
CQ3	Did something attack them in winter?
Sentence	Last summer the grass ate the koza in the field.
CQ1	Did something eat the grass in winter?
CQ2	Was the grass eaten?
CQ3	Did something eat the grass in the field?
Sentence	Weeks ago the cheese took the myska to its hole.
CQ1	Was the cheese taken by something?
CQ2	Did something take the cheese to the sofa?
CQ3	Did something take the cheese weeks ago?
Sentence	This morning them killed the flinta with a live round.
CQ1	Were they killed by a knife this morning?
CQ2	Did something kill them with a bullet this morning?
CQ3	Were they killed?
Sentence	Last month the kasu opened David with the key.
CQ1	Did David open something with a key?
CQ2	Was something opened with a crowbar?
CQ3	Was something opened last month?
Sentence	This morning the cleaner took the steku to work.
CQ1	Did she take something to her home?
CQ2	Did this happen this morning?
CQ3	Did she take something to her office?
Sentence	All morning the kocku rubbed she in the garden.
CQ1	Did she rub something in the garden?
CQ2	Did this happen in the living room?
CQ3	Did this happen at night?
Sentence	Last weekend the maltu mixed Brian with a stick.
CQ1	Was something mixed this morning?
CQ2	Did Brian mix something with a stick?
CQ3	Was something mixed by Brian?
Sentence	In the morning the ship kept the kotva in place.
CQ1	Did something hold the ship in place?
CQ2	Did this happen in the morning?
CQ3	Did the ship moved?
Sentence	This evening the krava fed its young with its milk.
CQ1	Did something feed its young with solid food?
CQ2	Did this happen in the evening?
CQ3	Did something feed its young?
Sentence	Last weekend the meat cut the dyka with its edge.

CQ1	Did something cut the meat?
CQ2	Was the meat cut yesterday?
CQ3	Did something cut the meat with its edge?
Sentence	Last year the delba carried water to his home.
CQ1	Was the water carried last year?
CQ2	Did something carry water to his home?
CQ3	Did the water move something?
Sentence	Last winter Sarah moved the sirku to the coal.
CQ1	Did she move something towards the coal?
CQ2	Was something carried away from the coal?
CQ3	Did this happen in winter?
Sentence	Today her rtenku applied she with a brush.
CQ1	Did she put something on with a pen?
CQ2	Was something applied with a brush today?
CQ3	Did she apply something?
Sentence	All evening the lecbu gave the nurse with a shot.
CQ1	Did the nurse give something?
CQ2	Was something given with a pill?
CQ3	Was something given by the nurse in the evening?
Sentence	Some time ago a studnu dug she at their home.
CQ1	Did she dig something?
CQ2	Did this happen at home?
CQ3	Did David dig something recently?

SPR Task: Testing Set

Sentences and Comprehension Question

Sentence	All night the pages printed the tiskna at work.
CQ	Did this happen at work?
Sentence	Today papers joined the seska in the office.
CQ	Were papers connected by something at home?
Sentence	All evening a strunu tied he to the present.
CQ	Was something tied to a tree?
Sentence	This morning the krida marked the pages with ink.
CQ	Did something mark the pages with paint?
Sentence	All day vegetables mixed the varka in the bowl.
CQ	Did James use something in the kitchen all day?
Sentence	All morning the maid used the prachu in the home.
CQ	Did this happen all evening?
Sentence	Today the lupu used a nurse at the clinic.
CQ	Did a nurse use something?
Sentence	Today the captain turned the kostku on the ship.
CQ	Did something sink the ship?
Sentence	Last weekend the fruit washed the prodna with water.

CQ	Did something weigh the fruit?
Sentence	Some time ago the brana blocked the road to town.
CQ	Did the brana open the road to town?
Sentence	Last weekend food burned the trouba in the kitchen.
CQ	Was the food burned today?
Sentence	All evening the hracka entertained them at home.
CQ	Were they entertained this morning?
Sentence	All year a paku built Peter with a board.
CQ	Did Peter build something?
Sentence	This morning the man changed the plenu in the bathroom.
CQ	Did she change something?
Sentence	Weeks ago the stuhu tied the man to the gift.
CQ	Did they tie something to the gift?
Sentence	Some time ago Mark bought an andku at the shop.
CQ	Did Mark sell something at the shop?
Sentence	In summer the liska ate the eggs at the farm.
CQ	Did this happen in winter?
Sentence	At night him welcomed the lestku with a cold drink.
CQ	Did this happen in the morning?
Sentence	Last winter the chuva cleaned the mess in the room.
CQ	Did this happen last winter?
Sentence	All week the car lent the punchna to the man.
CQ	Was the car lent to the man?
Sentence	Some time ago the lichtku gave he to his wife.
CQ	Did he give something to his wife?
Sentence	This evening she reduced the cenu at the shop.
CQ	Was something raised at the shop?
Sentence	All evening the smetnu mixed they in the bowl.
CQ	Did this happen at noon?
Sentence	In summer David cooked the kachnu with the fire.
CQ	Did he cook something in summer?
Sentence	This evening the room cleaned the sluzka with water.
CQ	Was the room cleaned by someone?
Sentence	Today the dilna delivered food to their home.
CQ	Was the food delivered?
Sentence	All weekend wood chopped the sekra in the forest.
CQ	Did something chop the wood in the forest?
Sentence	In the morning the bunda warmed him at the park.
CQ	Was he warmed by something in the morning?
Sentence	In the morning the kasnu built Katie with stone.
CQ	Did she build something?
Sentence	This evening Patrick wore a plastku in the rain.

CQ	Did he take off something in the rain?
Sentence	Last evening the propku pressed she on the paper.
CQ	Was something pressed against the wall?
Sentence	This morning she changed the ropu at the garage.
CQ	Was something changed by a man at the garage?
Sentence	In summer the zena bought the dress at the shop.
CQ	Did someone buy a dress?
Sentence	Last weekend the nuts ate the vevku in the tree.
CQ	Were the nuts eaten in the tree?
Sentence	In the morning the babka surprised them with cake.
CQ	Were they surprised with something in the morning?
Sentence	Last month the box sealed the paska in the bedroom.
CQ	Was something broken in the bedroom?
Sentence	In the evening a larku gave we to our friend.
CQ	Were we given something?
Sentence	Last summer Beth grew a papku in the garden.
CQ	Did Beth grow something on the windowsill?
Sentence	Last weekend a chalpu built she at the river.
CQ	Was something built by her at the river?
Sentence	This morning he heated the chatu with a fire.
CQ	Did he build a fire?
Sentence	Last night the skabka cleared the ice on the window.
CQ	Did this happen last night?
Sentence	Some time ago a kvetnu gave he to his wife.
CQ	Did this happen today?
Sentence	Some time ago they closed the mlecnu in the town.
CQ	Did they open something in the town?
Sentence	This evening she cleaned the bavnu with some water.
CQ	Did she burn something?
Sentence	At night her bag took the cizinka to the train.
CQ	Did she take her bag to the bus?
Sentence	Last year the cilka sharpened the knife on the stone.
CQ	Was a knife sharpened?
Sentence	Last weekend a ladku carried they to the room.
CQ	Did they bring something into the room?
Sentence	Last night a path cut the macka in the forest.
CQ	Did something happen in the forest?

Appendix H: Debriefing Questionnaire used in Pilot Studies 1 and 2, and Experiments 1 and 2

Candidate number________

Debriefing questionnaire

In the course of this experiment, you have indicated for every sentence whether you thought it belonged or did not belong to an artificial grammar system.

Reflecting on your performance, please estimate (in %) your **overall level of accuracy** in these judgements.

I estimate my overall level of accuracy to be …….. %.

While performing the tasks of the experiment, did you notice any particular rule or regularity? If yes, please indicate what you believe you have noticed.

As mentioned in the experiment, the sentences were not arbitrary. Reflecting now specifically on the **form of the foreign word** within the sentences, can you recall any specific rule or regularity?

Page 2: Personal information

1. What is your name?

2. Please indicate your age ……….., sex [m / f], and nationality ………..

3. What is your occupation?

If you are a student,

a) type of course (e.g., BA, PhD)

b) subject(s)

c) year

4. What is your language background? Please complete the following:

Native language(s):

Foreign language(s):

Appendix H: Debriefing Questionnaire Used in Experiment 3

Candidate number________

Debriefing questionnaire

While performing the tasks of the experiment, were you actively looking for any rules or regularities in the sentences?

What, if anything, did you notice about the sentences?

Reflecting now specifically on the **form of the foreign word** within the sentences, can you recall any specific rule or regularity?

Page 2: Personal information

1. Please indicate your age ……….., sex [m / f], and nationality ………..

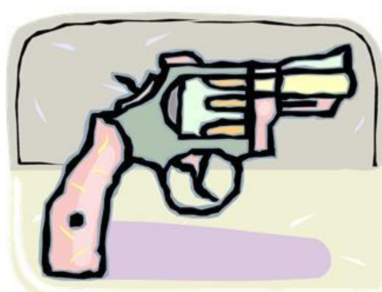
2. Course:

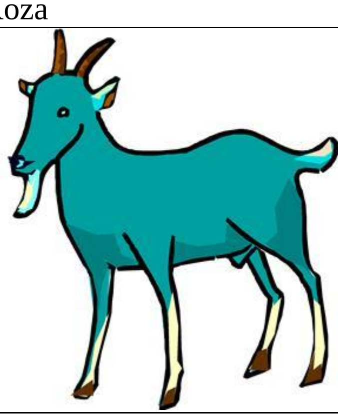
3. Year (e.g., 1st, 2nd, 3rd):

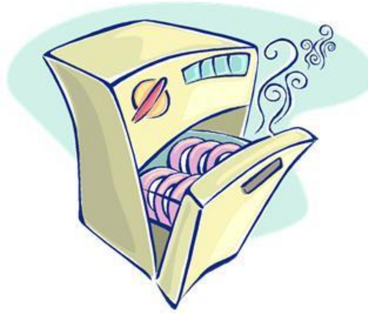
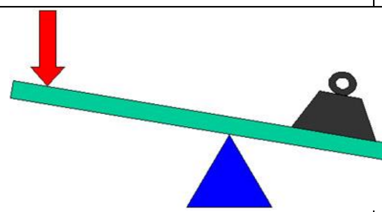
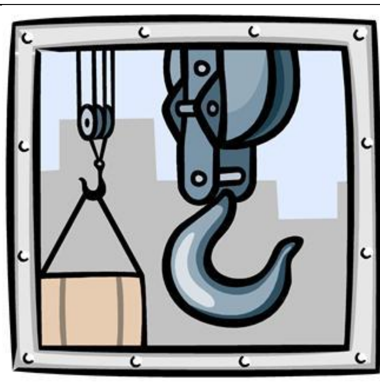
4. What is/are your native language (s)?

5. What foreign languages have you studied (please include level, e.g., Fluent, Beginner, GSCE, A-levels, etc.).

Appendix J: Pictures used in Pilot Studies 1 and 2, and Experiments 1 and 2

Baterka	Bedna	Brana
		
Brasna	Bylina	Britva
		
Dyka	Flinta	Flota
		

Guma	Hracka	Kasa
		
Kasna	Kotva	Kocka
		
Koza	Krava	Lecba
		
Lednicka	Lopata	Loutka
		

Malta	Mycka	Myska
		
Opona	Paka	Pekarna
		
Pilka	Pracka	Prodejna
		
Kladka	Reka	Salina
		

<table>
<tr><td>Salupa</td><td>Sirka</td><td>Skola</td></tr>
<tr><td></td><td></td><td>
</td></tr>
<tr><td>Sterka</td><td>Stetka</td><td>Studna</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>Susicka</td><td>Tovarna</td><td>Trouba</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>Vrtacka</td><td>Vyvrtka</td><td>Zahrada</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

Zehlicka	Zkouska	Zruda
		

Distractor Images		
		
		



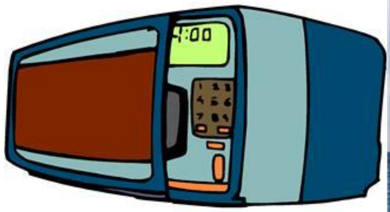


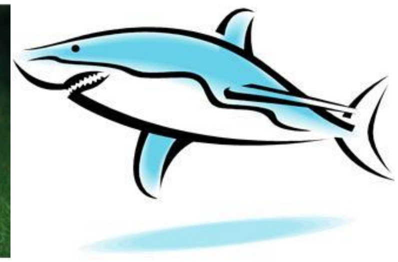


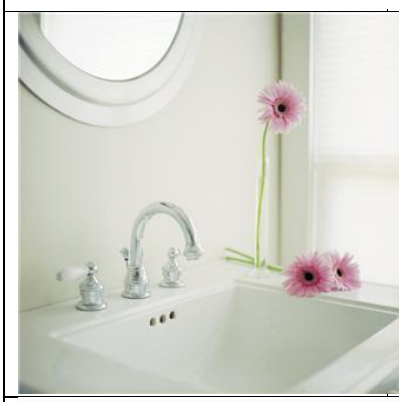




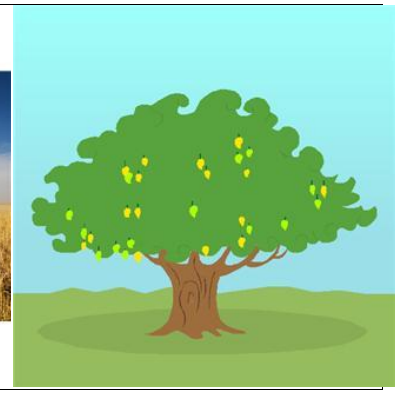
  


TELEPHONE

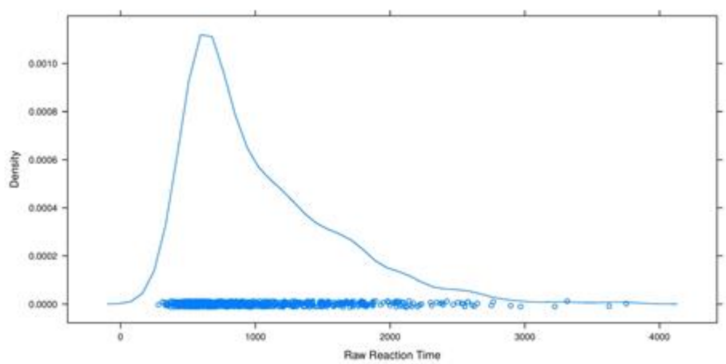



Appendix K: Supplementary Tables and Data from Experiment 3

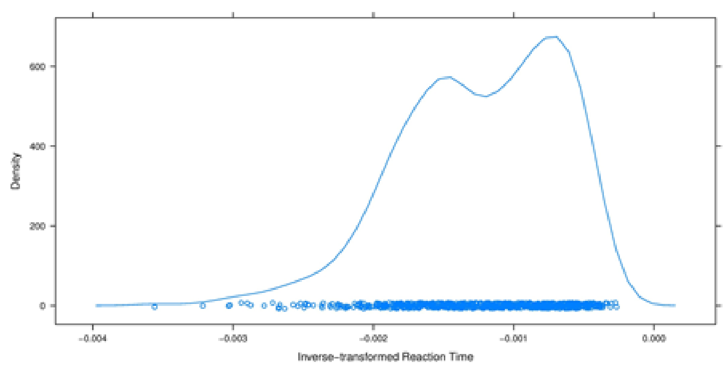
Contents of Appendix K

K1: Supplementary Data from the Critical Region of the SPR task

Untransformed Reaction Times



Inverse-transformed Reaction Times



Log-transformed Reaction Times

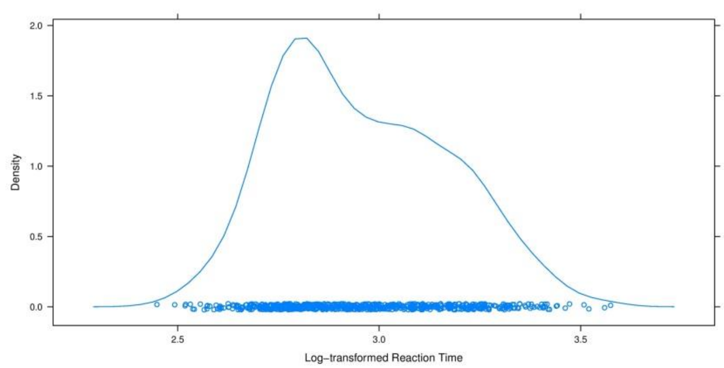


Figure A1. Density Plots of Raw, Inverse-transformed and Log-transformed Reaction Times of Critical Region Data from Experiment 3

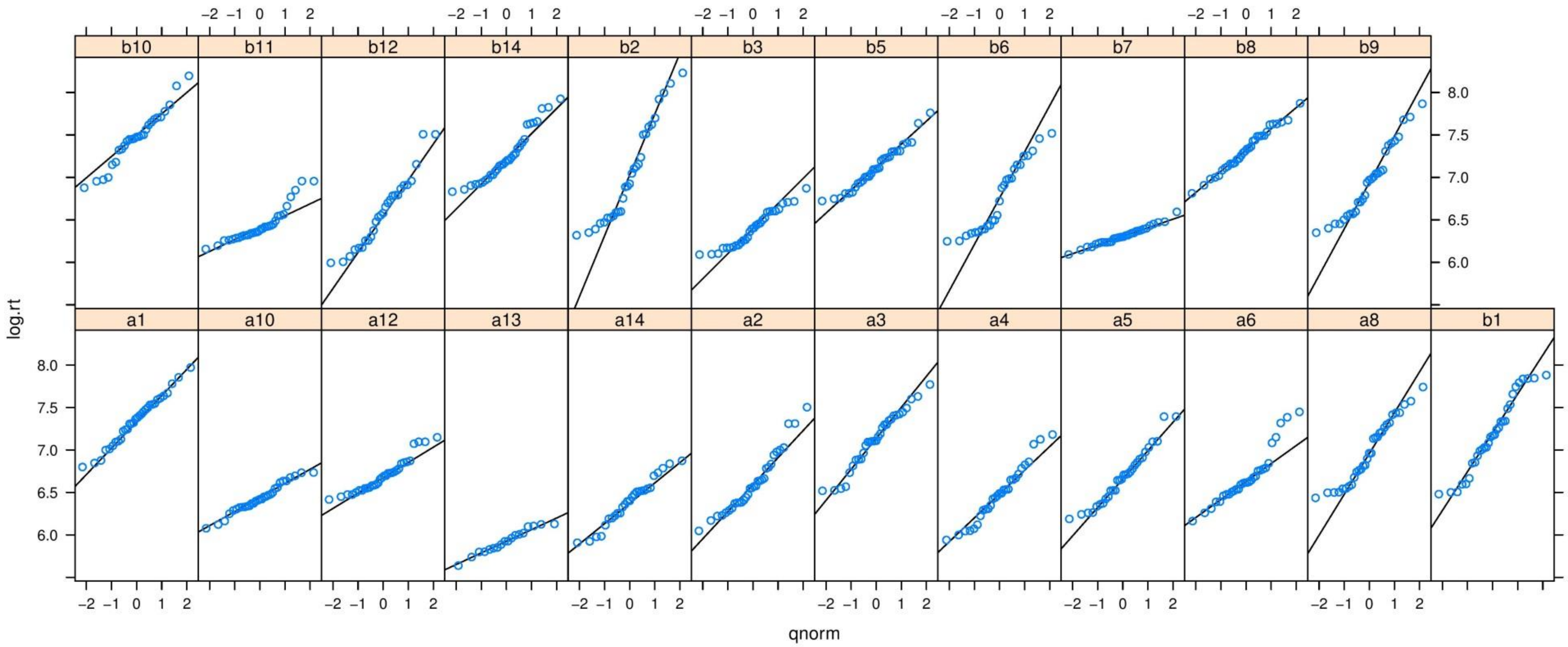


Figure A2. QQ plots of Log-transformed RTs of Critical Region Data by Participant

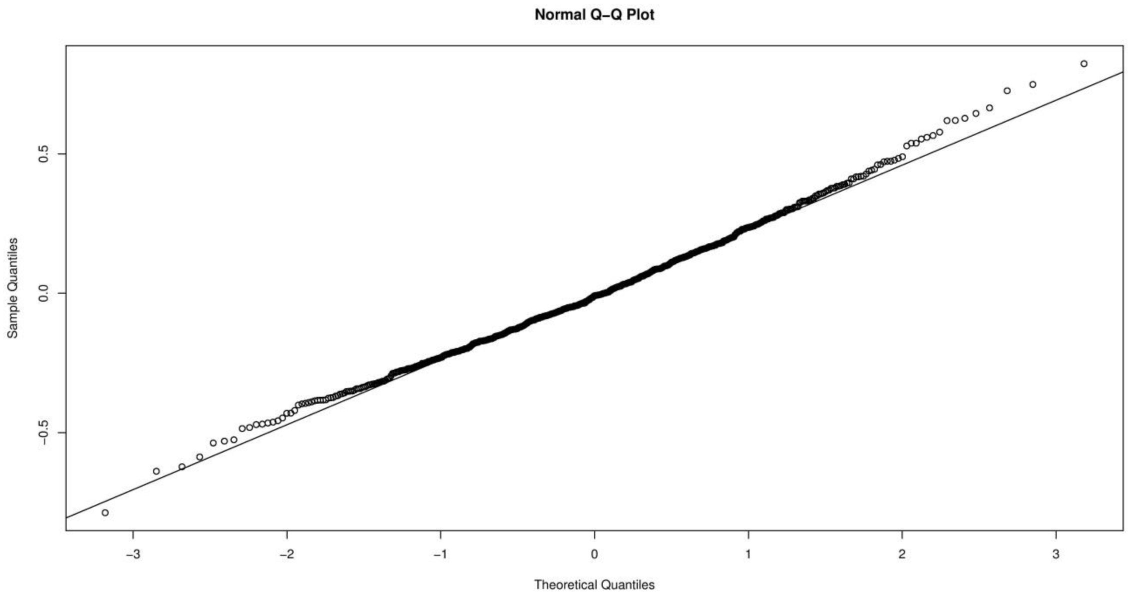


Figure A3. QQ plot of Residuals of the Best-fit Model for the Critical Region Data

K2: Supplementary Data from the Pre-critical Region

Table A1.
Average Response Time (ms) for Aware and Unaware Participants across Training and Testing Blocks for Segment 3

		Training Phase			Testing Phase		
Group		Block 1	Block 2	Block 3	Transition	Control	Violation
Aware	m	1178.66	1109.45	1128.92	1010.98	967.66	999.88
	sd	577.37	521.41	504.73	380.71	389.09	414.93
Unaware	m	820.75	772.46	734.25	687.37	712.15	685.12
	sd	528.34	406.19	397.78	394.25	507.84	386.92

Table A2.
Total Reaction Time (in ms) for All Aware and Unaware Participants for Control and Violation Items in Segment 3

All Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	831.01	824.88	837.13	818.39	843.62	807.65	854.36
	SE	23.92	33.30	34.43	34.14	33.58	31.11	36.34
	SD	458.92	451.65	467.02	463.05	455.46	421.95	492.98
Violation	M	806.61	808.74	804.50	814.78	798.49	764.26	849.20
	SE	23.79	28.49	38.13	28.77	37.92	28.28	38.05
	SD	455.68	385.43	517.24	389.17	514.33	384.38	514.71
Aware Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	1002.75	981.25	1019.73	1005.09	1000.41	987.78	1031.07
	SE	37.69	83.59	56.65	58.98	47.39	67.40	61.20
	SD	426.38	442.33	453.16	471.85	379.11	494.43	403.72
Violation	M	976.42	988.46	975.45	967.60	985.25	967.31	999.59
	SE	49.55	70.87	86.43	51.58	64.65	50.57	60.59
	SD	558.39	375.00	691.43	409.42	677.18	455.76	590.52
Unaware Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	728.74	723.08	734.41	713.48	744.00	705.89	751.59
	SE	28.61	40.66	40.42	38.72	42.24	35.24	45.14
	SD	443.23	445.42	442.83	424.20	462.76	386.02	494.46
Violation	M	727.34	736.03	718.66	739.80	714.88	688.08	766.61
	SE	23.75	33.52	33.78	32.64	34.61	30.78	393.83
	SD	367.94	367.14	370.08	357.54	379.15	337.18	35.95

Table A3.
Summary of Likelihood Ratio Tests in the Pre-critical Region

Fixed Effect	χ^2	df	p	r^2
Test version	4.01	5	.58	.51
Awareness	6.75	3	.08	.44
Case	0.58	4	.97	.44
Syntax	1.09	4	.89	.44
Locality	9.50	4	.05	.46
Grammaticality	5.17	4	.27	.46

Table A4.
Results from a Linear Mixed-effects Model Comparing Reaction Time (RT) across Control and Violation Items for Segment 3

	Fixed Effects			Random Effects	
				By subject	By item
Parameters	Estimate	*SE*	*t*	*SD*	*SD*
Intercept	2.91	0.05	58.47*	.11	.03
Aware	-0.02	0.06	-2.05*	–	.04
Locality	0.03	0.03	1.30	–	–
Aware: locality	-0.01	0.03	-0.15		

Note. All factors were coded using contrast coding, as follows: Syntax (−.5 = S-V-O, .5 = O-V-S), Locality (−.5 = near, .5 = far), Case (−.5 = Nominative, .5 = Accusative), Gram Model (−.5 = control item, .5 = violation item) formula: RT~aware*locality+ (locality|subject) + (aware|item).∗|t| > 2.0, indicating a significant effect (Gelman & Hill, 2007).

K3. Supplementary Data for Post-Critical Region

Table A5

Average Response Time (ms) for Aware and Unaware Participants across Training and Testing Blocks for Segment 5

		Training Phase			Testing Phase		
Group		Block 1	Block 2	Block 3	Transition	Control	Violation
Aware	m	1748.59	1701.65	1681.36	1488.55	1520.65	1499.42
	sd	1153.64	1145.54	1210.82	639.24	817.49	689.54
Unaware	m	1598.78	1634.17	1590.24	1509.30	1446.05	1452.59
	sd	1188.44	1118.57	1121.72	1094.55	1288.91	1126.31

Table A6.
Total Reaction Time (in ms) for All (Aware and Unaware) Participants for Control and Violation Items in Segment 5

All Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	1489.69	1478.36	1501.02	1483.64	1495.74	1497.13	1482.25
	SE	60.00	75.65	74.19	62.23	102.19	70.28	71.00
	SD	1150.96	1026.11	1006.39	844.08	1386.20	953.35	963.14
Violation	M	1445.63	1422.99	1470.27	1451.59	1441.67	1436.15	1467.11
	SE	52.01	60.40	84.84	78.75	68.17	78.45	68.52
	SD	997.79	819.31	1150.78	1068.22	924.66	1064.18	929.46
Aware Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	1520.65	1498.94	1542.36	1437.28	1604.02	1519.82	1521.48
	SE	72.54	109.26	96.31	82.49	118.63	112.87	91.64
	SD	691.44	874.05	770.51	659.88	949.02	902.99	733.13
Violation	M	1475.53	1415.14	1535.92	1427.72	1523.34	1400.64	1550.42
	SE	691.44	86.70	86.84	75.30	96.87	80.61	730.22
	SD	61.12	693.61	694.75	602.41	774.96	644.91	691.28
Unaware Participants		Total	Nom	Acc	SVO	OVS	Near	Far
Control	M	1473.34	1469.21	1477.47	1533.69	1412.99	1476.36	1470.33
	SE	63.13	100.19	77.28	84.55	93.76	89.05	89.82
	SD	978.02	1097.55	846.58	926.23	1027.04	975.44	983.91
Violation	M	1431.02	1416.51	1445.53	1468.98	1393.06	1452.43	1409.61
	SE	72.88	80.49	121.80	114.11	898.73	112.26	93.10
	SD	1129.00	881.74	1334.21	1250.00	82.04	1229.70	1019.91

Table A7.
Summary of Likelihood Ratio Tests in the Post-critical Region

Fixed Effect	χ^2	df	p	r^2
Test version	1.18	3	.76	.50
Awareness	0.60	1	.44	.51
Case	0.12	3	.99	.50
Syntax	5.57	3	.13	.51
Locality	2.13	3	.55	.51
Grammaticality	1.63	3	.65	.51
Spillover	0.21	1	.64	<.001