DEVELOPMENT OF CONSCIOUS KNOWLEDGE DURING EARLY INCIDENTAL LEARNING OF L2 SYNTAX

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy
in Linguistics

By

Phillip Hamrick, M.A.

Washington, DC
April 1, 2013
Second language acquisition (SLA) researchers have a long-standing interest in the effectiveness of providing learners with conscious knowledge (CK) through explicit instruction (e.g., Sanz & Morgan-Short, 2004); however, little is known about how CK develops under incidental learning conditions, without pedagogical intervention. In two experiments, the present dissertation addresses this gap by exploring the development of conscious knowledge of L2 syntax under incidental learning conditions, focusing on: (a) the mechanisms of L2 syntactic development and (b) the development of conscious knowledge of L2 syntax over time.

Experiment 1 addressed the mechanisms supporting L2 syntactic development under incidental learning conditions. Computational simulations of behavioral results indicated that Experimental participants who read sentences from a semiartificial language with probabilistic syntax in a moving-window paradigm learned via chunking mechanisms.

Experiment 2 extended these findings, focusing on the development of CK of L2 syntax over time. Following calls for triangulation in SLA research on awareness (e.g., Leow, 2000; Robinson, Mackey, Gass, & Schmidt, 2012), CK was operationalized using three measures of awareness: recognition memory (e.g., Perruchet, Bigand & Benoit-Gonin, 1997), retrospective verbal reports (e.g., Williams, 2005), and a novel subjective fluency judgment task. Eighty-two participants were randomly assigned to either an experimental or control group in one of three conditions, which differed in the amount of input. Participants were exposed to semiartificial language materials under the guise of a subjective fluency rating task. Afterward, participants
received an unexpected recognition memory test, in which they had to discriminate between old and new sentences. Finally, participants took part in a retrospective verbal interview.

Results indicated that learners’ CK was not equally captured by different measures of awareness, which underscores the need for methodologies that triangulate awareness using multiple measures. Participants formed CK in the form of declarative memories and verbalizable metalinguistic knowledge after as few as six exposures to new syntactic structures. Moreover, results indicated that L2 learners may rapidly and continuously acquire CK of L2 syntax even under incidental learning conditions, and is consistent with usage-based (e.g., N. Ellis, 2005; Robinson, 1996, 1997) and memory-based accounts of SLA (e.g., Paradis, 2009; Ullman, 2005).
ACKNOWLEDGMENTS

This dissertation would not have been possible without the input, support, and help from a great many people. Here I try to do some justice to their contributions, but I surely will underestimate them all, and for that I am sorry. All remain errors are my own.

My committee—Dr. Alison Mackey, Dr. Cristina Sanz, Dr. Patrick Rebuschat, and Dr. Rebecca Sachs—indulged me in undertaking the great risk that was this dissertation. Their brilliance, encouragement, and generosity kept me going when I thought I couldn’t. I sense that they all knew that saying some aspect of my work simply wasn’t good enough was a huge motivating force (not that they said it all that often). Each has been a mentor to me in different, but overlapping ways. I think Alison would say they were triangulating the best way to guide me. Alison took me under her wing despite my strong psycholinguistic persuasion, and taught me to love SLA and research methods. She also taught me that advisor and advisee do not have to have the same research interests or views, and even under such circumstances, she is still capable of being the best mentor (and promoter!) anyone could ask for. I can only hope to come close to that when I advise future students from a certain research area (you all know what that is).

Cristina was the first person to truly see the dissertation, since Experiment 1 was an offshoot of a study for her class. Since that class, I have felt as though I never had anything but her complete support for my research, even when she was criticizing it. I’m not sure how she managed that, but it is now something I aspire to do for my students, as I take it to be the hallmark of a truly great teacher and advisor. Patrick took me from being a graduate student who wanted to know about psycholinguistics to being, as one of my students put it, a Georgetown elder in psycholinguistics. I’ve been called many names in my day, but that one I really like. Finally, Rebecca Sachs deserves credit for two things: (1) being the only person who knows my work as well as I do as a result of endless hours and sessions of discussion and debate, and (2) teaching me that science, even the best and hardest of science, is still done by people, which means it could always be done better next time. For both of those things, my work is better than it could have possibly been without her, and for that I am truly grateful.
Endless thanks to all of my participants, in the pilot and main studies, and the teachers who let me recruit from their classes. This dissertation literally would not have happened without you all.

I also am indebted to a great many notable scholars, without whom I would have drowned in gutter water. The man who started me on the road to Linguistics is Dr. Salvatore Attardo, who taught me the meaning of intellectual pain. He got me interested in Linguistics back in 2006, and neither of us have let up since. Many wonderful and generous people have helped me on the way since then, and I am going to rattle them off in what I think is chronological order: Dr. Steven Brown, Dr. Vyv Evans, Dr. Ande Tyler, Dr. Jeff Connor-Linton, Dr. Robert DeKeyser, Dr. Michael Long, Dr. Lourdes Ortega, and two anonymous reviewers from *Language Learning*. Likewise, there are several people who have helped me with various aspects of this dissertation through detailed and extensive input at conferences, in personal meetings, and via email (or in some combination of the three): Dr. Nick Ellis, Dr. Jan Hulstijn, Dr. Ron Leow, Dr. Pierre Perruchet, Dr. David Shanks, and Dr. Michael Ullman.

Thanks to Erin Esch, who never failed to answer my questions, emails, more questions, more emails, handle forms, waivers, requests, orals setup, defense setup, and so on.

I would also like to thank my dear friends and colleagues in the Georgetown PsyLab. Back in 2009-2010, I started the PsyLab with Dr. Rebuschat, Julie Lake, Katie Kim, and Kaitlyn Tagarelli, each of whom have exerted considerable influence on my ideas and research in the last four years. I would also like to recognize the enormous contributions of PsyLab members who have since joined, Nick Pandža, Luke Amoroso, Dr. Rebecca Sachs (again), Mari Sakai, Haein Lauren Park, and Jessie Cox. Thank you all for listening to my really good ideas and giving me your time and input so that they could become really, really good ideas.

I also could not have completed this dissertation in the time I did without the help of Rachel Sawyer, my former LING-001 student and GUROP research assistant. Rachel’s intelligence, hard work, and reliability meant that she was able to collect data on days that I could not be at campus. I simply could not have asked for a better first research assistant.
Why do I do what I do? At the end of the day, I am always trying to figure out why kids can learn language so effortlessly and why it is that adults cannot do the same. A continual source of motivation and happiness for me throughout my time at Georgetown has been my nieces and nephew, all of whom I have got to see growing into language. Adanna, Cassidy, Chloe, Josephine, Olivia, and Weston: thank you for helping Uncle Phillip.

Second to last, but truly first, I want to thank my brilliant and beautiful wife, Dr. Jennifer Johnstone. We did it. You are my sweet, sweet someone coming through that door. You have listened to me become increasingly smart (and correspondingly insufferable) and you squinted and were patient with me while I said aloud to you everything I had to see to get here. It is physically impossible to thank you enough for that—or for the past eight years, so I’ll spend the rest of my years trying to. Normally, I say to you de nobis fabula narratur, but this time I thought of a better one, inspired by our play and Shirlington pub: Ah well, what matter, I say. Because of you, this will have been another happy day.

Finally, I would like to thank Georgetown University for these past four years of my life. You’ve made a happy man very old.

_Estragon_: I can’t go on like this.

_Vladimir_: That’s what you think.

- Samuel Beckett, _Waiting for Godot_

_This dissertation is dedicated to my parents._
# TABLE OF CONTENTS

Chapter I: Introduction ................................................................. 1
  1.1. Memory, consciousness, and (second) language acquisition .......... 1
  1.2. Key terms and definitions ..................................................... 5
    1.2.1. Consciousness .............................................................. 6
      1.2.1.1. Consciousness and awareness ...................................... 6
      1.2.1.2. Conscious knowledge ............................................... 7
      1.2.1.3. Lower-order explicit knowledge .................................. 7
      1.2.1.4. Higher-order explicit knowledge .................................. 9
      1.2.1.5. Implicit knowledge ................................................... 10
      1.2.1.6. Incidental and intentional learning ............................... 11
      1.2.1.7. Implicit and explicit learning .......................... 12
    1.2.2. Memory ............................................................... 12
      1.2.2.1. Declarative memory ............................................... 12
      1.2.2.2. Recognition memory ............................................. 13
    1.2.3. Knowledge representation ............................................. 18
      1.2.3.1. Abstract rule ...................................................... 18
      1.2.3.2. Microrule ............................................................. 19
      1.2.3.3. Statistical computation .......................................... 20
      1.2.3.4. Chunk ................................................................. 21
Chapter II: Literature Review .......................................................... 24
  2.1. Assumptions ........................................................................... 24
    2.1.1. Theoretical approaches implicating consciousness in language learning ........ 24
    2.1.2. Memory-based approaches to second language acquisition .................. 29
      2.1.2.1. Declarative and procedural memory models ....................... 29
      2.1.2.2. Usage-based and exemplar models ................................ 32
    2.1.3. Three observations ...................................................... 34
  2.2. Most of language acquisition is incidental ................................ 35
  2.3. Incidental learning generally results in a knowledge base
      that is largely available to consciousness ................................... 35
Chapter I: Introduction

1.1. Memory, consciousness, and (second) language acquisition

Until recently, it was considered improbable that memory (e.g., Chomsky, 1965) or consciousness (e.g., Jackendoff, 1987) played any central role in the acquisition of natural language syntax.\(^1\) This makes intuitive sense, since syntax allows humans to generate an infinite number of new utterances (for which they could not have memory) using the systematic properties of a language (of which people are not conscious). However, there is an increasing number of arguments that memory and consciousness do in fact play some necessary roles in language. Indeed, researchers working in a variety of paradigms now argue that syntactic information is stored in memory (e.g., Culicover & Jackendoff, 2006; Goldberg, 1995, 2006; Hagoort, 2005; Jackendoff, 2011; Tomasello, 2003; Ullman, 2004, 2005) and that consciousness is crucial for narrowing the scope of processing to manageable fragments of experience through chunk formation and the development of metalinguistic knowledge (e.g., N. Ellis, 2005; Perruchet & Vinter, 2002; Schmidt, 1990, 1994, 1995, 2001). Empirical evidence consistent with these claims comes from research in cognitive neuroscience (e.g., Ettlinger, Novis, Wang, & Wong, under review; Morgan-Short, Finger, Grey, & Ullman, 2012; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Novick, Kim, & Trueswell, 2003; for overviews, see Paradis, 2004, 2009; Ullman, 2001, 2004, 2005), usage-based accounts of child language (L1) acquisition (e.g., Abbot-Smith & Tomasello, 2006; Kidd, Lieven, & Tomasello, 2010; Kidd & Lum, 2008;

---

\(^1\) Traditionally, generative approaches assume that syntax acquisition is a modular process and, as such, does not rely on other cognitive capacities such as consciousness and memory (e.g., Fodor, 1983; Schwartz, 1999). Moreover, in more recent approaches, clear roles for long-term memory are left out altogether (e.g., Hauser, Chomsky, and Fitch, 2002), making it unclear what role memory is given in current theory (see also, Jackendoff, 2011).
Stoll, Abbot-Smith, Lieven, 2009; for overviews, see Tomasello, 2003), analyses from some generative approaches to linguistics (e.g., Culicover & Jackendoff, 2006; Pollard & Sag, 1994), second language (L2) acquisition (e.g., DeKeyser, 1997; Rebuschat, 2008; Rebuschat & Williams, 2009, 2012; Robinson, 1996, 1997, 2005, 2010) and computational psycholinguistics (e.g., Bod, 2006, 2009; Perruchet & Vinter, 1998; Perruchet & Peereman, 2004; Perruchet & Tillman, 2010; Reitter, Keller, & Moore, 2011; Vosse & Kempen, 2000).

Crucially, these approaches agree that syntactic development is, at least in its earliest phases, based on lexical memory: we acquire conscious fragments of lexical knowledge, which have both exemplar-specific and abstract properties and which tacitly encode syntactic information (e.g., walk + ed, I want some X; Culicover & Jackendoff, 2006; N. Ellis, 2005; Goldberg, 2006; Tomasello, 2003). This conscious knowledge is stored as complex forms (i.e., chunks) in the neural substrates supporting declarative memory (e.g., Bod, 2009; Ullman, 2001, 2004, 2005). These chunks then act as “concrete seeds” from which the “abstract trees” of syntax subsequently develop (N. Ellis, 2005, p. 320).

Renewed interest in memory and consciousness in recent accounts of language acquisition echoes interest in the roles of memory and consciousness in learning in other fields, such as second language acquisition (SLA). SLA researchers have an ongoing interest in the roles of memory (e.g., DeKeyser, 1998, 2007; Morgan-Short et al., 2012a,b; Paradis, 2004, 2009; Randall, 2007; Ullman, 2004, 2005) and consciousness (e.g., N. Ellis, 2005; Krashen, 1981, 1985; Leow, 1999, 2000; Robinson, 1995; Schmidt, 1990, 1994, 1995, 2001; Williams, 1999), especially in instructed contexts, since memory and consciousness may influence L2 development (Schmidt, 1995). Indeed, the notion that conscious knowledge may influence L2 development (possibly in some positive way) has underscored nearly thirty years of research in
explicit instruction (for overviews, see, e.g., Alsadhan, 2011; Goo, Granena, Novella, & Yilmaz, 2009, forthcoming; Norris & Ortega, 2000; Spada & Tomita, 2010). However, most of this SLA research has been aimed at the use of interventions or experimental manipulations that either provide learners conscious knowledge or direct their attention in such a way to promote learner awareness of certain types of forms (e.g., Long & Robinson, 1998). Little effort has been directed at understanding how learners develop conscious knowledge without being prompted to do so (although see Park, 2011). However, it is well-established empirically that even under incidental learning conditions, people become spontaneously aware of structural regularities in training stimuli without being asked to do so and without intending to do so (e.g., Haider & Frensch, 2005; Rose, Haider, & Buchel, 2010; Wagner, Gais, Haider, Verleger, & Born, 2004). Perhaps the most common finding in the incidental learning literature is that learning under incidental conditions typically results in a knowledge base which is largely accessible to consciousness, often consisting of declarative memories (as measured by recognition tasks, e.g., Perruchet, Bigand, & Benoit-Gonin, 1997; Shanks & Johnstone, 1999; Shanks & Perruchet, 2002; Tunney & Shanks, 2003; Wilkinson & Shanks, 2004) and/or verbalizable knowledge (e.g., Dulany, Carlson, & Dewey, 1985; Hama & Leow, 2010; Hamrick & Rebuschat, 2011, 2012, 2013; Perruchet & Pacteau, 1990; Rebuschat, 2008; Rebuschat, Hamrick, Sachs, Ziegler, & Riestenberg, 2013; Rebuschat & Williams, 2012; for overviews, see Perruchet, 2008; Rebuschat, forthcoming; Shanks, 2005).

The present dissertation brings together these various lines of research to address a gap in the literature which is derived from four observations.

---

2 Throughout the dissertation, I consider the incidental learning and implicit learning literatures as essentially isomorphic. Although there are differences between the learning phenomena (e.g., the implicit learning literature typically measures awareness, etc.) and the fields that focus on them (e.g., SLA has a strong focus on incidental learning, regardless of the issue of awareness, e.g., Hulstijn, 2003; Pellicer-Sanchez & Schmitt, 2010) the learning conditions in both research paradigms are typically incidental in the sense defined later in this chapter.
1. SLA researchers have a vested interest in issues of consciousness and memory, which have hitherto primarily been investigated through explicit instruction or pedagogical treatments.

2. Incidental learning research (primarily from the implicit learning literature) has shown that the development of conscious knowledge (in the form of metaknowledge and declarative memories) is a common result of incidental learning.

3. The bulk of naturalistic L1 and L2 acquisition is incidental (VanPatten & Williams, 2007).

4. Declarative memory and consciousness may play important roles in language acquisition, including early L1/L2 syntactic development.

From these observations we can deduce that the incidental acquisition (3) of conscious knowledge (2) should be of intrinsic interest for SLA researchers, for whom the role of conscious knowledge in L2 development is a central question (1 and 4). The gap is that despite being a potentially valuable source of information for SLA researchers, surprisingly little is known about the development of conscious L2 knowledge during the course of incidental learning (as opposed to the development of conscious L2 knowledge from, say, pedagogical interventions or strategic, intentional learning). There are at least two plausible reasons for this gap: First, the focus of instructed SLA research has been on optimizing instructional interventions. This emphasis, by definition, reduces the amount of interest in incidental learning conditions. Second, researchers interested in incidental learning have typically been interested in implicit cognition. Although many of these researchers report evidence of the development of conscious knowledge under

---

3 Here I am referring to learners who seek to develop their L2 to levels of very high proficiency and often engage in non-intentional, immersion-type L2 learning activities. This observation excludes many cases of L2 acquisition. For example, many students take foreign language classes for academic or employment requirements, for recreation, etc. This observation also runs the risk of the monolingual bias (Ortega, 2009).
incidental learning conditions, their focus has been on determining the nature of implicit knowledge, which is more difficult to establish. Consequently, the literatures have left the issue of the incidental learning of conscious knowledge relatively unaddressed. The present dissertation aims to address this gap in terms of the development of conscious knowledge during the incidental learning of L2 syntax, following on recent work on the role of memory (e.g., Culicover & Jackendoff, 2006; Hagoort, 2005; Robinson, 1996, 1997; Ullman, 2004, 2005) and consciousness (e.g., N. Ellis, 2005; Perruchet & Vinter, 1998, 2002; Schmidt, 1990, 1994, 1995, 2001) in early syntactic development.

The aim of the next chapter is to clarify and substantiate the above points in more detail. First, I will review the central findings from incidental learning research, pointing out that the common findings there are consistent with memory-based approaches to language. I will then review SLA research pertaining to the role of conscious knowledge in L2 development, suggesting that there is a generally positive, albeit complex relationship between conscious knowledge and L2 development. Finally, this will lead me to explicate the gap in our knowledge of the present research questions. Before describing the gap, I will first provide an overview of the key terms and definitions that are used in this dissertation.

1.2. Key terms and definitions

Terms relating to learning, consciousness, and memory are notoriously difficult to use precisely and consistently. Differences in the use of terms and definitions may come from researchers’ different backgrounds (e.g., philosophy v. neuroscience) and emphases on different phenomena (e.g., instructed SLA v. memory), and this can lead to confusion (for example, see the recent debate on the implicit/explicit interface between N. Ellis (2005) and Paradis (2009), which
centers largely on possible definitions of the term *interface*). Consequently, it is important to lay out clear and detailed definitions for key terms that are used in this dissertation. Therefore, the goal of this section is to provide as much clarity as possible in my operationalization of terms in this dissertation. In what follows, I give definitions for key terms, focusing first on terms relating to conscious knowledge and implicit knowledge since they are terms that apply to subsequent concepts. I then proceed to terms relating to declarative memory, and finishing by defining terms related to knowledge representation (e.g., rules, chunks, etc.).

1.2.1. Consciousness

1.2.1.1. Consciousness and awareness

I use the terms consciousness and awareness interchangeably to refer to phenomenal, subjective experience. I generally use them in the sense that Searle (1997) uses the term consciousness:

“Consciousness consists of those states of sentience, or feeling, or awareness, which begin in the morning when we wake from a dreamless sleep and continue throughout the day until we fall into a coma or die or fall asleep again or otherwise become unconscious.”

In this case, the objects of awareness can be perceptual (e.g., sensation of heat), a cognitive representation (e.g., semantic knowledge of the hot thing you are touching, a memory of camping on your seventh birthday), or a higher-order mental state or metaknowledge (e.g., knowing that you are having an experience of the sensation of heat). This definition is broad, maybe too much so to generate falsifiable predictions; however, it is used because of its breadth in allowing us to further break down consciousness and awareness into levels. Accordingly, consciousness can be thought of as

---

4 In this dissertation, I do not discuss conscious states such as vivid dreams, since they are not relevant for SLA.
having levels depending on the object of consciousness.\(^5\) For example, perceiving the form of the word \textit{jumped} and having a corresponding cognitive representation of its meaning active in consciousness is, to some degree, different from directing a higher-order state at that meaning (a metacognitive state, i.e., I know that I am thinking about the meaning of \textit{jumped}) and is also different from having higher-order knowledge about the morphophonological rules that govern the combination of \textit{jump} and \textit{–ed} (entertaining a metalinguistic thought).

\textbf{1.2.1.2. Conscious knowledge}

Conscious knowledge is a general term used to refer to both lower-order and higher-order explicit knowledge. In other words, it is used to refer to situations in which people may have either lower-order or higher-order explicit knowledge or both.

\textbf{1.2.1.3. Lower-order explicit knowledge}

Whenever a percept (e.g., hearing a word) or cognitive representation (e.g., conceptual-semantic knowledge of the meaning of the word) enters into consciousness, the basic conscious experience of it is considered to be lower-order awareness, or lower-order explicit knowledge.\(^6\) This can be exemplified by noticing a tree. I define “noticing” in the technical sense, following Schmidt (2001), as the conscious subjective experience of “surface structure of utterances in the input, instances of language, rather than any abstract rules or principles of which such instances may be

\footnotesize
\(^5\) I am not the first to make this point. Leow (1997) distinguishes between levels of awareness in accordance with, essentially, what the contents of awareness are.

\(^6\) Percepts do not have to have meaning (e.g., perceiving the color orange, seeing a printed word in an unknown foreign language) and cognitive representations do not have to have corresponding percepts (e.g., understanding the meaning of the word \textit{freedom}).
exemplars” (p. 5). Perception of a tree is followed by the activation of knowledge about trees, which guides your processing of the tree as a tree (Perruchet & Vinter, 1998, 2002). The unconscious attention and memory mechanisms active in this process give rise to lower-order explicit knowledge: your phenomenal experience of the tree. Note that this can occur—indeed, does occur most of the time—without a corresponding higher mental state or thought directed at it (i.e., I am seeing this tree right now and I know that is what I’m doing). In language, lower-order explicit knowledge is active when you hear [ðə kæt] (the cat), and it consists of the perception of a sequence of phones or syllables and meaning as a unitary conscious experience. At this level of consciousness, you are not aware of the implicit mechanisms guiding the phrase structure rules or phonotactics of the cat, although they are operating. Thus, the development of lower-order explicit knowledge is an iterative process. Perception of an external stimulus forms a cognitive representation of that stimulus in memory, which subsequently guides processing of that stimulus in the future.

When lower-order explicit knowledge has been acquired conscious experience is changed. For example, imagine that you are in the checkout line at the grocery store. In the next line down, you hear two people having a conversation in a language you do not know. What does their conversation sound like to you? You probably hear syllables or combinations of syllables as units that probably do not match the actual syllable combinations that make words in the foreign language (i.e., the chunks you hear probably are not initially the chunks that are actual words in the foreign language). Now imagine that you hear a sound sequence that sounds like computer. What happens? Chances are that the stream of syllables is interrupted, and you have the

---

7 Detection refers to the cognitive registration of a stimulus and does not require awareness (Tomlin & Villa, 1994). Detection acts as a filter or selector of a subset of what enters awareness. What is detected is typically, but not necessarily always, isomorphic with the contents of lower-order conscious experience (Perruchet & Vinter, 2002).
conscious experience of hearing the word computer, not as a random set of syllables, but as a whole, meaningful unit. This is because you have active lower-order explicit knowledge of the word computer. This basic experience, so fundamental and automatic in everyday life, is very much conscious. You are not aware of all of the associative and neural processes that gave rise to this experience. Rather, these implicit processes have shaped cognition to give you the conscious experience of hearing this word as a whole meaningful unit.

Although I return to this point in more detail later (section 2.1.1), it is worth noting here that from a theoretical standpoint, lower-order phenomenal experience and lower-order explicit knowledge are thought to provide a natural filter for information. In other words, any given instance of phenomenal awareness is narrow in scope, presumably due to the central role played by attention in constraining awareness (e.g., Cohen, Cavanagh, Chun, & Nakayama, 2012; Perruchet & Vinter, 2002; Schmidt, 1995, 2001; Tomlin & Villa, 1994). It is this iterative property of lower-order experience that gives it power during learning: forming representations from conscious experience which guide further conscious experience, leading to increasingly complex representations (e.g., Tye, 1995).

1.2.1.4. Higher-order explicit knowledge

Higher-order explicit knowledge refers to metaknowledge that is available to consciousness. In the domain of language, higher-order explicit knowledge is prototypically, but not mandatorily, metalinguistic (e.g., I know that the regular rule for the past tense in English is to add –ed to the main verb). Indeed, within SLA, higher-order explicit knowledge is often thought of and/or researched as metalinguistic knowledge. For example, the notion of awareness at the level of

8 Consider a similar effect in text: pabikugolatimubanacomputerbunamikopabola. If you know the word computer, then you will cease seeing ngrams of syllables and instead perceive it as a whole unit.
understanding (Leow, 1997; Schmidt, 2001) is subsumed under the term higher-order explicit knowledge. Crucially, higher-order explicit knowledge, as metaknowledge, can be learned independently of the linguistic knowledge to which it refers. For example, people can learn the metalinguistic rule for the English past tense before they ever hear any actual exemplars of it. Higher-order explicit knowledge need not be strongly metalinguistic in the traditional sense. Consider the learner who develops the following generalization about English: “‘the’ comes before words like dog and car, but not people.” This hypothetical example includes higher-order explicit knowledge without metalinguistic terminology, and does so in a way that is seemingly more tied to lexical items. Such examples are included in the definition of higher-order explicit knowledge in order to account for the fact that people can have such knowledge without having the formal training that allows them to make more complex metalinguistic judgments.

1.2.1.5. Implicit knowledge

Within the implicit learning research tradition, implicit learning has typically been operationalized as follows: When people improve their performance on a task without being conscious of what has been learned from the training material, then they have acquired implicit knowledge (cf. Perruchet, 2008, p. 610). The very possibility of implicit knowledge or implicit memory has been debated, although this largely stems from arguments over the use of the term knowledge or the very idea of dividing knowledge and/or memory on consciousness, not over the existence of implicit aspects of the mind and brain (e.g., Perruchet & Vinter, 1998, 2002; Shanks & Berry, 2012). For example, some would consider lower-order explicit knowledge to be implicit (e.g., Dienes & Perner, 1999). Therefore, terminological differences are probably due to different research goals in the study of consciousness. With this in mind, implicit knowledge is
defined as a given state of mechanisms and processes whose operations are intrinsically unavailable to consciousness, although their products may be. For example, recall the cat as discussed above. Implicit mechanisms govern the underlying statistical regularities or rules embodied in this noun phrase, although the products of these mechanisms (i.e., the sound and meaning of the cat) constitute a conscious experience. Another way to conceive of implicit knowledge that is consistent with this admittedly loose definition is as the connection weights and algorithms in a connectionist neural network, although this concept is itself not without substantial problems (Perruchet & Vinter, 2002, p. 369).

1.2.1.6. Incidental and intentional learning

In non-experimental contexts, incidental learning is generally defined as learning that proceeds without the intent to learn the knowledge they actually do learn (e.g., just using Spanish and incidentally learning morphosyntax; or trying to learn Spanish words and incidentally learning morphosyntax), and intentional learning as learning that proceeds with participants’ intent to learn the knowledge that they actually do learn (e.g., I’m going to sit down and learn Spanish morphosyntax). In terms of research design, I follow Hulstijn (2003): incidental learning conditions refer to learning experiments in which participants are not informed that they will be tested. Moreover, for present purposes, incidental learning conditions refer to experimental manipulations in which learners are deliberately misled regarding the purpose of the training task. This is done to prevent, as much as possible, participants from strategic learning processes toward the target forms. Intentional learning conditions refer to learning experiments in which participants are forewarned of a test phase and are instructed to learn the relevant forms.
1.2.1.7. Implicit and explicit learning

Implicit and explicit learning are defined, briefly, as follows: Implicit learning is the acquisition of implicit knowledge under incidental learning conditions. Explicit learning is the acquisition of conscious knowledge under intentional learning conditions. Following R. Ellis (2005; 2009, p. 6) and Schmidt (1994, p. 20), it is important to keep in mind that the learning processes of implicit and explicit learning are distinct from the conscious status of the resulting knowledge. Since intentionality in the learning phase does not equate with the conscious status of the resulting knowledge used in the test phase. As such, implicit knowledge may well be acquired under intentional learning conditions, and conscious knowledge may be acquired under incidental learning conditions. Indeed, the latter is the very subject of this dissertation. However, there may be strong objections to labeling the incidental acquisition of conscious knowledge as implicit learning, and the intentional learning of implicit knowledge as implicit learning. Therefore, when I refer to these I will use the unfortunately cumbersome labels incidental learning that results in conscious knowledge and intentional learning that results in implicit knowledge.

1.2.2. Memory

1.2.2.1. Declarative memory

In a descriptive sense, I use the term declarative memory to refer to memory for personal (i.e., autobiographical) events, semantic information, and other facts that are, at least in part, available to consciousness (e.g., Berry, Shanks, Li, Rains, & Henson, 2010; Norman, 2002; Shanks & Berry, 2012; Ullman, 2004, 2005). Declarative memory also supports implicit knowledge, such as knowledge of syntax in lexicalist and usage-based accounts (e.g., N. Ellis, 2005; Hagoort, 2005) and implicit knowledge of complex, but abstract morphosyntactic structures in cognitive
neuroscience accounts (e.g., Marslen-Wilson & Tyler, 2007). Declarative memory consists of at least two subtypes, *episodic memory* and *semantic memory*. Episodic memory is memory for personal events, while semantic memory is memory for facts and knowledge about the world. Declarative memory structures are critical for the binding of the contents of consciousness into new representations which are input from other cortical regions. General principles of associative learning appear to govern this binding process (e.g., Cohen, Poldrack, & Eichenbaum, 1997; McClelland et al., 1995; Perruchet & Vinter, 1998, 2002; Ullman, 2004). Declarative memory abilities are generally associated with activity in the hippocampus and medial temporal lobe, but declarative memories are eventually represented independently of these regions and become more reliant on neocortical areas (see Ullman, 2004, 2005, for overviews of the other neural regions associated with the declarative memory system). In terms of language, declarative memory is thought to subsume lexical or vocabulary information (Paradis, 2009; Ullman, 2004, 2005), chunks (Ullman, 2004, 2005), metalinguistic abilities (DeKeyser, 2007; Paradis, 2004, 2009; Ullman, 2004), and other analogically-derived complex structures (Ettlinger, Wang, Novis, Wong, & Patrick, under review). Moreover, declarative memory is thought to underlie early phases of syntactic development in both children and adults via the memorization of complex forms (Ullman, 2005).

**1.2.2.2. Recognition memory**

Declarative memory also underlies recognition memory judgments (Medina, 2008). Recognition memory is the ability to accurately assess whether or not a stimulus has been encountered before. There are competing accounts of recognition memory, all dealing largely with the contributions of *recollection*, *familiarity*, and *priming* or *fluency*. Recollection is the ability to judge whether a stimulus has been seen before or not by retrieving the original
instance. Familiarity is the conscious experience that a stimulus has been experienced before without recalling further information. Priming is when participants use (implicitly) their own processing speed as a heuristic for judging whether a stimulus is old or new. There is fairly substantial evidence that recollection and familiarity represent independent cognitive and neural processes (e.g., Yonelinas, 2002). On the other hand, there is an increasing consensus that familiarity and priming may be supported by the same underlying cognitive and neural mechanisms (e.g., Berry, Shanks, & Henson, 2008; see Medina, 2008, for an overview). Some, too, have suggested that complete independence between priming and recollection may not be warranted (Kinder & Shanks, 2003).

Regardless of whether a single-, dual-, or multiple-mechanism account proves appropriate for modeling recognition memory, there are some general points of agreement between models. For instance, models of recognition memory tend to consider familiarity to be best modeled as a single strength-of-evidence memory process, much like that in a signal detection theory (SDT) approach (Figure 1 below). To illustrate the SDT approach, let us assume that judgments in a recognition memory task are made on the basis of a familiarity signal. In the recognition memory test, participants are asked to judge whether stimuli are old (previously encountered) or new (not previously encountered). To make a familiarity-based recognition memory judgment, the familiarity mechanism registers the similarity between a stimulus and a memory is assessed along a hypothetical “strength-of-evidence” scale (represented as the X axis in Figure 1 below). It is assumed that, due to prior exposure, old stimuli should have higher probability of being judged similar by the familiarity mechanism (represented as the Y axis in

---

9 Although some argue that familiarity is unconscious (Scott & Dienes, 2008), I do not subscribe to this view since (1) the feeling of familiarity is conscious and (2) it is this conscious feeling that we act on when making a recognition memory decision. The knowledge base giving rise to feelings of familiarity may be implicit (hence, it is familiar rather than recognized), but familiarity itself is a conscious, subjective feeling, and is arguably based on declarative memory (Squire & Zola, 1996).
Figure 1 below) It is also assumed that the familiarity mechanism has a threshold, past which the mechanism “recognizes” a stimulus as being old. This threshold is represented by $C$ in Figure 1. Stimuli falling to the left of $C$ are new items correctly judged to be new (correct rejections) or old items incorrectly judged to be new (misses). Stimuli falling to the right of $C$ are old items correctly judged to be old (hits) or new items incorrectly judged to be old (false alarms). Despite its apparent simplicity, the SDT model has had substantial success in accounting for a wide range of recognition memory findings (Medina, 2008).

Figure 1.1. A signal detection theory (SDT) model of old-new recognition memory judgments (taken from Berry, Shanks, & Henson, 2008).

![Signal Detection Theory Model](image)

However, participants need not use recollection or familiarity when judging the old-new status of stimuli in a recognition memory task. Instead, participants can use priming as the basis for their judgment. In this case, participants are no longer using a conscious knowledge base, but, rather, an implicit one. Consequently, there has been criticism that objective tests of conscious knowledge such as recognition memory may be contaminated by implicit knowledge (i.e.,
recognition memory tests may not exclusively be tapping into conscious knowledge sources, Reingold & Merikle, 1990). Shanks and Johnstone (1999, p. 1436) point out that empirical evidence often indicates priming in the presence of conscious knowledge (e.g., Perruchet & Amorim, 1992), which can be taken as evidence that tests of both implicit and conscious knowledge are in fact measuring implicit knowledge.

Why is priming held to be implicit? Well, despite the fact that priming often occurs in the presence of recognition memory, it also often occurs in the absence of recognition memory (making it likely to be implicit). For example, Leung and Williams (2011) trained participants on a semiartificial language in a reaction time task. The training phase consisted of 82 exposures to a regular pattern (based on agent-patient and adult-child distinctions in artificial determiners). This was followed by a testing phase that systematically violated the regular determiner patterns. Reaction times for these violation sentences were significantly slower than reaction times for non-violation sentences, indicating that participants had learned something about the pattern. Importantly, 80% of the participants reported not being aware of the underlying pattern, suggesting that the basis of the priming effect in participants’ reaction times was implicit. It is worth noting that similar results have been reported in experiments using the serial reaction time task (e.g., Destrebecqz & Cleeremans, 2001; Reed & Johnson, 1994); however, many of these studies have been unable to be replicated (see Shanks, 2005, for an overview). Thus, there is some evidence that implicit knowledge in the form of speeded response (priming) may influence recognition memory judgment, and, as such, needs to be accounted for in any model of recognition memory.

However, more recently, several researchers have presented evidence that priming may not be due to a pure implicit knowledge base, and have argued that familiarity can also
adequately account for priming effects in recognition memory tasks. Shanks and colleagues (Berry, Shanks, & Henson, 2008; Berry, Shanks, Speekenbrink, & Henson, 2012; Shanks & Berry, 2012; Shanks & Perruchet, 2002) have repeatedly demonstrated that a formal model with a single memory variable, *familiarity*, when exposed to different amounts of noise associated with recognition memory tasks and reaction time tasks, can produce dissociations between recognition and priming found in amnesic individuals as well as in cognitively normal adults. Similar findings using other methodologies have been obtained elsewhere (see Medina, 2008, for an overview). Work in the dual-mechanism approach has also converged on the importance of ideas from SDT. Some researching the dual-mechanism approach have put forward the idea that recognition is driven by two separate processes, recollection and familiarity, each of which consists of a strength-of-evidence memory variable and a threshold point, past which stimuli are recognized as having been previously seen (Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). In this account, priming can also be the result of the familiarity-based mechanism. Taken together, recent work is converging on the notion that familiarity and priming may index a common underlying memory system that is neither purely implicit nor purely conscious in the traditional sense (Shanks & Berry, 2012). In this dissertation, I do not attempt to distinguish different models of recognition memory. However, priming is taken as an index of implicit knowledge that can contaminate conscious knowledge measures, such as recognition memory. This is *not* due to any fundamental disagreement with current approaches that unite familiarity- and priming-based sources of knowledge. Rather, taking priming as a measure of implicit knowledge provides a sort of methodological safety net. It provides one more source of information about what was being measured in the recognition memory test in Experiment 2. These issues are revisited in chapters 4 and 5.
1.2.3. Knowledge representation

1.2.3.1. Abstract rule

Abstract rules are mental operations that combine or transform other representations of variables (Marcus, 2001). Rules are conceived of as symbolic or algebraic categorical structures that are independent of the frequencies and probabilities of surface structure. For example, the past tense rule in English applies to regular verbs no matter how frequently or infrequently those verbs occur. No matter how similar a structure is to a correct rule, if it breaks a rule in any way, it is not grammatical. As Manza and Reber (1997, p. 75) put it, the representation of rule knowledge contains

[Little, if any, information pertaining to specific stimulus features; the emphasis is on structural relationships among stimuli. The key here is the notion that the mental content consists, not of the representation of specific physical forms, but of abstract representations of those forms.

Crucially, there is some debate over whether abstract rules as implicit knowledge exist. In principle, abstract rules may be implicit knowledge structures governing combinatorial symbol manipulation (e.g., Chomsky, 1965; Marcus, 2001), combinations of syntactic templates (e.g., as in Merge, Chomsky, 1995; Unification-based grammars, Culicover & Jackendoff, 2006; Hagoort, 2005) or production sequences of chunks (e.g., Anderson & Lebiere, 1998).

Alternatively, abstract rules may only be conscious knowledge used, for example, for reasoning and conscious problem-solving (e.g., Cleeremans & Destrebecqz, 2005; Hulstijn, 2002).\textsuperscript{10}

In Experiment 1, abstract rule knowledge is operationalized as categorically accurate performance on a given grammaticality judgment test structure. In Experiment 2, conscious knowledge...
abstract rule knowledge is operationalized as complete metalinguistic knowledge about one or more syntactic structures used in the experiment.

1.2.3.2. Microrule

Microrules (Dulany et al., 1984; Dulany, Carlson, & Dewey, 1985) refer to conscious, “folk” heuristics that learners may develop during the course of learning. These may be different for each learner, but they generally consist of “conscious rules, each of limited scope, and many of imperfect validity” (Dulany et al., 1984, p. 541)\(^{11}\). Learners often develop microrules regardless of whether they are instructed to find rules or not (e.g., Mathews et al., 1989; Pothos, 2007). An example in natural language is an ESL learner having formed the following conscious rule: “In English I cannot place *an* in front of a mass noun (e.g., *a rice*).” There are undoubtedly many instances where this simple rule works, but also plenty of cases where it does not (e.g., for delineated units such as water bottles: *a water*, as in “Could you hand me a water?”). Thus, this microrule is limited in scope (does not specify information about *the* or count nouns) and is invalid in certain circumstances. Crucially, microrules are, by definition, conscious. They are also similar to conscious abstract rules, with the primary difference being that microrules are imperfect, incomplete, and/or partially-inaccurate. Within the context of this dissertation, microrules are operationalized as any partially-accurate metalinguistic knowledge participants develop during the course of the experiment.

1.2.3.3. Statistical computation

\(^{11}\) Note here that the distinction between microrule and conscious abstract rule is best conceived of as a continuum. Microrules are thought of as imperfect and/or limited, whereas conscious abstract rules may be thought of as more robust: more accurate and applying to a wider range of instances.
What statistical computation entails is a matter of debate. Some researchers posit that human cognition is endowed with the ability to unconsciously make powerful, complex statistical inferences, the type of which a statistician would make (e.g., N. Ellis, 2002, 2005, 2006; Gopnik, Wellman, Gelman, & Meltzoff, 2010; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). Computational models operating under this assumption (e.g., Bayesian models) have been able to capture a variety of human behaviors (e.g., induction of syntax, reasoning, decision-making) but have been criticized for their lack of constraints and psychological plausibility (e.g., Altmann, 2010; Perruchet & Poulin-Charronnat, forthcoming). More psychologically (if not neurally) plausible approaches suggest that implicit statistical computations can be thought of as unconscious mental processes that track predictive dependencies or as learning probabilistic cue-outcome relationships (i.e., contingency learning). Connectionist networks provide one way of simulating these latter sorts of computations (e.g., Misyak, Christiansen, & Tomblin, 2010; Shanks, 1995; Williams, 2009). For example, a simple recurrent network (SRN; Elman, 1990) operates by predicting what item comes next in a sequence, and learns on the basis of its prediction accuracy or error. Williams (2010; Williams & Kuribara, 2008) has demonstrated that this model is able to capture some aspects of L2 syntactic development. In his simulations of syntax learning experiments on human adults, Williams trained the SRN to predict syntactic category sequences. To do this, the SRN was given incremental input. At each time step, the SRN would be given a syntactic category (e.g., NP) and the SRN would produce a prediction about what syntactic category would come next (e.g., VP). When the SRN made a correct prediction, it strengthened the connection weights that led to the correct prediction, but when the SRN made an incorrect prediction, it weakened the connection weights that led to the incorrect prediction. After enough training, the network was able to successfully predict syntactic category
sequences on the basis of the probabilities between syntactic categories. It has been demonstrated elsewhere that this learning mechanism makes the SRN very good at capturing transitional conditional probabilities, e.g., the likelihood that one item will follow another in given sequence (Elman, 1990; Perruchet & Peereman, 2004). More details on the computational architecture of the SRN are found in the section on computational models in Chapter 3.

In this dissertation, statistical computation refers to processes that track predictive dependencies like that found in an SRN. This definition is preferred over others primarily because of their psychological plausibility and because SRNs have been the most frequently-used models of statistical learning and language acquisition to date. In this dissertation, the existence of statistical computations is established in two ways: First, through on-line reaction time measures in Experiment 1. These data are analyzed to see if reaction times are faster at points of higher transitional probability than at points of lower transitional probability, consistent with what is expected by theories of statistical learning based on computation (e.g., Hunt & Aslin, 2001). Second, evidence of statistical computations is achieved by seeing how well an SRN can simulate the syntax learning results of human adults in Experiment 1. If the SRN simulates human performance, then there is evidence for some functionally similar mechanisms in the human mind.

1.2.3.4. Chunk

A chunk is any mental unit of language which encodes sequential information (e.g., in the sequence A-B-C-D, BC encodes sequential information: B comes before C and C comes after B). Given the present focus on syntax, I apply this definition specifically to sequential information
regarding word order, syntactic category order\textsuperscript{12}, and larger clausal sequences. On this view, a whole memorized sentence may serve as a chunk (e.g., I want to eat), as does a fragment of a sentence (e.g., I want to), a partially schematic representation of a sentence (e.g., I want to X), and a fully abstract construction (e.g., NP VP NP). Furthermore, humans appear to encode sequential information from the position of single items (e.g., in the sequence A-B-C-D, B is the second item; Knowlton & Squire, 1994, 1996; Schuck, Gaschler, & Frensch, 2012; Schuck, Gaschler, Keisler, & Frensch, 2012). Such positional chunks are conceivable in natural language, as in slot and frame patterns (e.g., Pine & Lieven, 1997). For example, the preposition in encodes sequential information (in is typically phrase-initial). Anchor chunks, which occur at the edges of sequences need only consist of a single item (e.g., ABCD), but may consist of more items (e.g., ABCD).

In principle, chunks can be conscious, implicit, or both (this will be elaborated further in section 2.1.1). As noted earlier, when processed, the cat contains conscious content (e.g., the spoken and/or written form and conceptual meaning of the word), along with content that is typically not conscious (e.g., phrase structure, phonotactics). Moreover, chunks can be more or less abstract depending on their lexical-specificity, which is consistent with the item-based learning approach found in usage-based linguistics (Tomasello, 2003), cognitive linguistics (e.g., Evans, 2009), and construction grammar (e.g., Goldberg, 2006). For instance, a child may acquire want some X, where the chunk is partially lexically-specified (e.g., contains two words and one abstract slot for words fitting a certain syntactic category or thematic role)\textsuperscript{13}. By

\textsuperscript{12} This also applies to related semantic categories such as thematic role order in argument structure.

\textsuperscript{13} If chunks are abstract, then they can be filled in with concrete chunks (i.e., words and morphemes). In this case words can embed themselves into the slots of an abstract chunk, and in general, this is the same principle as placing words or other chunks into chunks recursively. Note that such recursion is fully consistent with the chunk-based associative learning approach to language provided we posit another capacity for holding multiple chunks together.
adulthood, chunks can be fully abstract syntactic templates (Hagoort, 2005). Thus, chunks can be, in theory, abstract (Perruchet & Vinter, 2002). By abstract, I mean surface-independence in the following sense: “is verb-ing” can be considered a chunk (like a slot and frame pattern) consisting of lexically and morphophonologically specific information (is and –ing) with a slot that is equivalent to the lexical category VERB. This slot is abstract in the sense that it is not tied to any verb in particular, but rather applies to a whole class of lexical items.14 Regardless of how lexically-specified they are, chunks probably reside in the declarative memory system and are hierarchically organized with superordinate and subordinate levels (N. Ellis, 1998, 2005; Ettlinger et al., under review; Servan-Schreiber & Anderson, 1990; Skehan, 1998, Chapter 2; Ullman, 2004, 2005; Wray & Perkins, 2000).

In this dissertation, chunk knowledge is assessed by comparing the results of human performance with that of a computational model whose only learning mechanism is chunk formation (PARSER; Perruchet & Vinter, 1998). Details on this procedure are provided in chapter 3.

in conjunction (such as working memory, Perruchet & Poulin-Charronnat, forthcoming; Rey, Perruchet, & Fagot, 2012; Merge, Chomsky, 1995; Unification, Hagoort, 2005)).

14 It is especially important here to note that the issue of abstractness or surface-independence is not contested by different approaches to incidental learning. Rather, what is contested is the issue of the mental computations or processes operating on representations, be they abstract rules, statistics, or associative processes (Perruchet & Poulin-Charronnat, forthcoming).
Chapter II: Literature Review

2.1. Assumptions

Recent research in SLA, psycholinguistics, and cognitive neuroscience has provided increasing evidence that consciousness and memory play important roles in language development. While there are various theories, models, hypotheses, and frameworks associated with this view, there is a small subset of approaches that needs elaboration before continuing. In this section, I briefly review theories dealing with consciousness and memory in second language acquisition and in the psychology of learning. I begin by discussing theoretical approaches to consciousness and memory in SLA and cognitive science. I then review research that illuminates the gap in the research based on the observations outlined in Chapter 1. Then, I conclude by describing how these issues motivate this study and outlining the research questions addressed in Experiments 1 and 2.

2.1.1. Theoretical approaches implicating consciousness in language learning

Issues of consciousness bear on a variety of issues in language learning and are of interest to language teachers and theorists alike (Schmidt, 1995). Since the early 1980s (e.g., Krashen, 1981) language teachers and instructed SLA researchers have maintained a strong interest in the role of consciousness in L2 development; however, the role of consciousness in theories of language learning is by no means a settled matter. Some ascribe little or no role to consciousness (lower-order or higher-order) especially in the development of syntax or other aspects of grammar (e.g., Hauser et al., 2002; Jackendoff, 1987; Schwartz, 1999; VanPatten, 2011) . For
example, VanPatten (2011) takes the position that syntax “is not an aspect of language that is learned in the way that learning is traditionally understood. Learning in the traditional sense involves the interaction of learner-directed attention to particular features of an environment with general learning mechanisms...In fact, syntax is not learned at all. Syntax is derived from the interaction of environmental data (input) and UG (plus the mechanisms that interface UG with the environment)...This does not mean that learners do not consciously pay attention when listening to someone else speak or somehow use explicit processes to learn vocabulary, morphology, and other aspects of language. The point here is that what learners do explicitly has little to do with how syntax grows." For VanPatten, the contrast is between input + UG and input + general learning mechanisms. Importantly, lower- and higher-order consciousness and memory are intrinsic components of general learning mechanisms. Taking this to its logical end supposes that VanPatten will not ascribe any causal role in the development of syntax to consciousness or memory-based learning mechanisms.

VanPatten’s view contrasts with that taken in this dissertation and by others, who, for example, give more weight to lower-order explicit knowledge. Indeed, a variety of positions on the matter can be found in the SLA literature on the implicit/explicit interface. For example, the strong interface position holds that conscious knowledge can be “converted” into implicit knowledge (e.g., DeKeyser, 2003, 2007). If correct, then this means that conscious knowledge could quite directly impact the development of implicit syntactic knowledge. There are also a variety of weak interface positions. Some posit an influence of conscious knowledge on implicit knowledge when learners are developmentally ready (e.g., Pienemann, 1989). Others are consistent with the position that the contents of conscious experience are operated on by implicit knowledge when

15 Note that DeKeyser’s strong interface view is only a technicality. In his view, conscious knowledge can become so automatized so as to be functionally indistinguishable from implicit knowledge. This does not entail that conscious knowledge actually becomes implicit knowledge, but it does obviate the need to posit it.
learning mechanisms (e.g., N. Ellis, 2005; R. Ellis, 1993, 1994; Perruchet, 2005; Robinson, 1996, 1997). Still, others propose that conscious knowledge supports learner output, which then is processed by implicit learning mechanisms (e.g., Schmidt & Frota, 1986; Sharwood Smith, 1981). Finally, there are non-interface positions, in which conscious knowledge exerts no direct influence on the development of implicit knowledge (e.g., Paradis, 2004, 2009; Ullman, 2004, 2005). However, there is room in these positions for an indirect influence, where conscious knowledge supports usage until implicit knowledge can do the work of processing the input. Finally, some posit a role for lower-order explicit knowledge throughout the entire span of language acquisition (as opposed to being just a catalyst for implicit knowledge). For instance, Perruchet and Vinter (2002) posit an iterative cycle between lower-order awareness and learning.

What fills the contents of consciousness is learned and what is learned alters subsequent conscious experience, e.g., if you learn that *dog* is a word, your subsequent conscious experience is altered such that you come to perceive it as a word.

Rather than diminishing a role for consciousness, these approaches posit that lower-order explicit knowledge provides fundamental organizational constraints to the input. As N. Ellis (2005, p. 312) points out: “Compared to the vast number of unconscious neural processes happening in any given moment, conscious capacity evidences a very narrow bottleneck. However, the narrow limits of consciousness have a compensating advantage: Consciousness seems to act as a gateway, creating access to essentially any part of the nervous system.” Applied to lower-order explicit knowledge, consciousness provides our implicit learning mechanisms with a narrow scope for operation, reducing the computational workload on the brain and preventing problems of combinatorial explosion (e.g., Perruchet & Poulin-Charronnat, forthcoming). These approaches to consciousness are consistent with Schmidt’s Noticing
Hypothesis (1990, 1994, 1995, 2001), which argues that *noticing* (i.e., lower-order awareness) is crucial for language learning. Schmidt (2001, p. 5) defines *noticing* as the conscious subjective experience of “the surface structure of utterances in the input, instances of language, rather than any abstract rules or principles of which such instances may be exemplars.” In terms of the Self-Organizing Consciousness approach (Perruchet & Vinter, 2002) and Implicit Tallying Hypothesis, *noticing* provides natural limits on the subjective contents of consciousness, which guides implicit learning mechanisms (N. Ellis, 2005, 2006; Perruchet & Poulin-Charronnat, forthcoming), which must be given a manageable amount of input or otherwise suffer catastrophic failure.

Such lower-order conscious experience is central to another theory of consciousness and language learning, embodied in the computational model PARSER and expanded in the theory *self-organizing consciousness* (SOC; Perruchet & Vinter, 1998, 2002). In PARSER and the SOC, the content of conscious experience (i.e., what is noticed) forms a chunk. The contents of a chunk are available to awareness, but tacit structure of the chunk is encoded implicitly by associative learning mechanisms. That is, the conscious contents of experience form a chunk and the features and properties of this chunk are encoded and tallied by implicit associative mechanisms (N. Ellis, 1998, 2005). Thus, as noted earlier, chunks are two-sided, containing both conscious and unconscious aspects of mind. So, what is noticed forms a chunk. If that chunk is encountered again in the input, a match occurs between the contents of the current conscious experience and prior experience, and the chunk is strengthened through repetition. If the chunk is not encountered again, it is forgotten due to decay and interference (e.g., from other, similar input).
PARSER and the SOC assume that all adult learning begins with conscious experience. Noticing results in the formation of conscious chunks, and exposure to more input results in the formation and accumulation of more complex chunks (a similar view is espoused in the Fundamental Similarity Hypothesis, Robinson, 1996, 1997). Chunks constrain the problem space for implicit mechanisms, preventing combinatorial explosion. Implicit mechanisms iteratively operate on these chunks, resulting in changes to that lower-order explicit knowledge and subsequent conscious experience (for more detail, see section 2.3.1.4). Given this premise, Perruchet and Vinter (2002) argue that phenomena that appear to require abstract, unconscious rule processes, may be due to the interplay between complex conscious knowledge and relatively simple, implicit associative processes.

Not all approaches that favor a role for conscious knowledge in learning posit such a strong and pervasive role for it as is found in PARSER and the SOC. For instance, N. Ellis’ (2002, 2005) Implicit Tallying Hypothesis gives a central role to lower-order explicit knowledge in creating a stimulus representation. However, once that representation “is firmly in existence…need never be noticed again; yet as long as it is attended\textsuperscript{16} to for use in the processing of future input for meaning, its strength will be incremented and its associations will be tallied and implicitly cataloged” (N. Ellis, 2002, p. 174). In other words, lower-order explicit knowledge\textsuperscript{17} is required for early learning, but not for the subsequent tuning of the language system. Other approaches emphasize higher-order explicit knowledge. For example, DeKeyser’s (2007) skill acquisition approach to SLA focuses on the necessity of higher-order explicit knowledge in early learning. This type of knowledge, regardless of how it is learned, is necessary

\textsuperscript{16} It is unclear whether “attended” in this view leads to a phenomenal, lower-order awareness, but, as attention is possible without awareness, it is certainly not a mandatory interpretation.

\textsuperscript{17} Higher-order explicit knowledge may also effect such implicit tallying (N. Ellis, 2005).
to support extensive practice. Over time, the practiced knowledge is automatized, making the retrieval of higher-order explicit knowledge unnecessary for language use.

In sum, there are a number of theoretical approaches to consciousness and language learning, only a few of which were surveyed here. These approaches posit that conscious knowledge (lower-order explicit knowledge, higher-order explicit knowledge, or both) are required for at least the early phases of learning. This point of view underlies many of the predictions and hypotheses made in the present dissertation. However, these theoretical approaches do diverge in the extent to which they implicate consciousness over the extent of language learning. Some posit a mandatory role for the initial development of lower-order explicit knowledge (e.g., Perruchet & Vinter, 2002; Schmidt, 1990, 1995), and some invoke a facilitative and/or mandatory role for both lower- and higher-order explicit knowledge in early L2 development (e.g., DeKeyser, 2003, 2007; N. Ellis, 2005). In the general discussion in chapter 5, I will return to these theories in order to assess the extent to which they shed light on the results of the experiments in chapters 3 and 4.

2.1.2. Memory-based approaches to second language acquisition

In this section, I briefly review two general approaches to language acquisition that give prominent roles to memory in syntactic development: Declarative and Procedural memory models and Usage-based/Exemplar models.

2.1.2.1. Declarative and procedural memory models

The Declarative/Procedural (DP) model (Ullman, 2001, 2004, 2005) posits that language acquisition and processing depend on two distinct long-term memory systems: declarative
memory and procedural memory. According to the DP model, declarative memory, which was described in chapter 1, supports the acquisition and storage of the components of the mental lexicon (i.e., conceptual meanings, phonological information, and grammatical information). Declarative memory supports early phases of syntactic development through the memorization of complex forms (e.g., lower-order explicit knowledge of chunks) and through metalinguistic knowledge (e.g., higher-order explicit knowledge of language rules). Procedural memory, which is posited to rely on distinct neural substrates (e.g., the basal ganglia) is posited to underlie combinatorial aspects of more mature grammar. However, learning takes longer in this system, so early L2 syntactic development is posited to rely more on declarative memory until the analogous (but anatomically separate and computationally distinct) knowledge in procedural memory can develop. The same is true, “even in very young children learning their native language, [where] complex forms as well as idiosyncratic knowledge are predicted to be memorized in declarative memory before grammatical rules are abstracted in procedural memory (Ullman, 2005, p. 163). For adults, however, the DP model posits that the combination of heightened declarative learning abilities and the mild attenuation of the procedural memory system lead language learners to have a more pronounced and enduring reliance on declarative memory for grammar. Finally, the DP model assumes that procedural memory is an implicit memory system, while declarative memory contains both implicit and conscious memories.

Paradis (2004, 2009) also invokes the declarative/procedural divide in accounting for language acquisition. Like Ullman’s DP model, Paradis’ approach posits that adult L2 learners will have to rely heavily on declarative memory in acquiring grammar. However, Paradis assumes that the declarative/procedural memory distinction is isomorphic with the explicit/implicit memory distinction, a view which is increasingly contested (e.g., Reder, Park, &
Kieffaber, 2009; Shanks & Berry, 2012). Moreover, because Paradis assumes isomorphism between declarative/procedural and explicit/implicit, he also assumes that non-conscious aspects of vocabulary (e.g., that obey takes direct object) are a part of the procedural memory system\(^\text{18}\). However, more research is needed to determine whether the neural bases of the grammatical aspects of the lexicon are associated with declarative or procedural memory systems.

Now, both of these memory models broadly agree in their central tenets (e.g., that acquired grammar is handled by procedural memory, while words and other idiosyncratic, memorized forms reside in declarative memory). Although they subtly disagree on several points, they both produce the same broad prediction for early phases of syntactic development; namely, they both predict that declarative memory for chunks and exemplars are first steps in the acquisition of syntax (e.g., Paradis, 2009, p. 93; Ullman, 2004, 2005). This prediction is central to the hypotheses made later in experiments 1 and 2.

Finally, it is worthy of note that there are other memory-based models that do not rely on the declarative/procedural distinction. For example, on the basis of a wide range of neuroimaging data, Hagoort’s (2005) Memory, Unification, and Control model considers syntax to be stored in declarative (lexical) memory in the form of abstract, syntactic templates (or chunks), and there are no syntactic rules that add any syntactic elements beyond what are stored in memory (i.e., no rules that add nodes or unpronounced copies during parsing). Likewise, computational models developed by Bod (2009), Reitter et al., (2011), and Vosse and Kempen (2000) all assume that syntactic knowledge is stored as hierarchically organized abstract chunks in the lexicon. All of these models have had success accounting for human behavioral phenomena in syntactic processing. In short, memory-based models that do not focus on the declarative/procedural

\(^{18}\)This view is inconsistent with some empirical data. For example, Hamrick and Rebuschat (2013) found evidence that both implicit and explicit lexical knowledge are supported by the same underlying mechanisms, suggesting that a single memory system is responsible for both aspects of the lexicon.
distinction place the bulk of syntax into the lexicon. These approaches are consistent with recent theoretical work within usage-based linguistics (e.g., N. Ellis, 2005; Goldberg, 1996; Langacker, 1987; Lakoff, 1987; Tomasello, 2003), to which we turn next, and is increasingly popular in some recent non-transformational generative approaches as well (e.g., Chomsky, 1995, 1998; Culicover & Jackendoff, 2006; Pollard & Sag, 1994).

2.1.2.2. Usage-based and exemplar models

Like the declarative/procedural models, usage-based approaches to SLA (e.g., Cadierno, 2008; N. Ellis, 2002, 2005, 2006, 2008; Ellis & Cadierno, 2009; Ellis & Larsen-Freeman, 2009; MacWhinney, 2008; Robinson & N. Ellis, 2008; Roehr, 2008) assume that early syntactic development is based on memory for exemplars or fragments of exemplars (i.e., chunks; N. Ellis, 1998, 2005), although this is rarely discussed overtly in terms of memory systems (e.g., Abbot-Smith & Tomasello, 2003; Bybee, 2006; Goldberg, 2006; Tomasello, 2003). Consider an example from the child language acquisition literature. Cameron-Faulkner, Lieven, and Tomasello (2003) conducted a corpus analysis of the child-directed speech of 12 mothers speaking English to their 2-year-old children. They found that 45% of all maternal utterances began with a set of just 17 words (e.g., what, that, it, you, come, etc.), while over 50% began with 52 frames (e.g., Can you…, Look at…, What did…, etc.). Children’s utterances correlated strongly with the frequencies of these frames (i.e., what the mother frequently said, the child frequently said). This general finding has been replicated in other studies and in other languages (e.g., N. Ellis, 2013; Hamrick, ms; Rowland, Pine, Lieven, & Theakston, 2003; Stoll, Abbot-Smith, & Lieven, 2009). In short, it appears that early syntax acquisition is shaped by frequent exemplars, a position consistent with a role for memory in early syntactic development.
Because usage-based approaches rarely make specific predictions regarding memory systems in the brain, it is not always easy to tell whether or not they diverge from declarative/procedural models in their predictions. One possible point of divergence concerns abstraction of syntactic regularities from exemplars. Some approaches posit that abstraction derives, not from a distinct syntax module, but rather from stored exemplars. Thus, abstract syntax is regarded as not rigidly distinct from the lexicon (e.g., Bod, 2009; N. Ellis, 2013; Goldberg, 2006; Tomasello, 2003, pp. 297-299). Therefore, this account does not posit a need for a separate storage system for syntax, e.g., a procedural memory, and indeed de-emphasizes the need for extensive complex combinatorial operations (although it does not explicitly negate them). Usage-based approaches are, in principle, compatible with multiple-mechanism accounts like the declarative/procedural models. The compatibility comes from the usage-based assumption that language is acquired, not through language-specific syntax-learning mechanisms, but through general cognitive capacities. This view is certainly consistent with Ullman’s DP model, which emphasizes the domain-general nature of the memory mechanisms supporting language. For the purposes of the present dissertation, which is focused on early syntactic development, whether or not syntax ultimately comes to rely on a separate memory system is immaterial. Both approaches make the same prediction regarding early syntactic development, which is our present concern: early syntactic development is driven by memory for exemplars and chunks.\(^\text{19}\)

In sum, memory-, usage-, and exemplar-based approaches to language, while differing in many underlying assumptions and broad theoretical predictions, tend to converge on the view that early syntactic development is supported by declarative memory for exemplars, chunks, and

\(^{19}\) There are opposing views, e.g., Lidz, Waxman, and Freedman (2003), who ascribe mature syntactic competence to infants in the earliest phases of development.
high frequency word-slot patterns. Ullman’s DP model (2004, 2005; Morgan-Short & Ullman, 2012) predicts this to be the general pattern, regardless of whether we are speaking of child language acquisition or L2 learning. Likewise, usage-based and exemplar approaches believe L2 acquisition to be fundamentally similar to first language acquisition, and so predict similar reliance on memory for exemplars (e.g., N. Ellis, 2005; Robinson, 1996, 1997). The prediction that early syntactic development will be based, at least in part, on memory for exemplars or chunks will be maintained in this dissertation.

2.1.3. Three observations

In addition to assuming that the earliest phases of syntactic development involve conscious knowledge and memory-based mechanisms, I also make three other observations. These observations lead ultimately to a gap in our current knowledge, which is that we know little about the incidental learning of conscious knowledge of L2 syntax.

- **Observation #1**: Most naturalistic L2 learning is incidental (e.g., VanPatten & Williams, 2007).

- **Observation #2**: Incidental learning generally results in a knowledge base that is largely available to consciousness (e.g., Perruchet, 2008; Shanks, 2005).

- **Observation #3**: Conscious L2 knowledge is an established point of interest for SLA research (e.g., Schmidt, 1995, p. 2).

In the following three sections (2.2., 2.3., and 2.4.), I review the evidence for each of these observations.
2.2. Most of language acquisition is incidental

The first observation is a logical one. There is simply not enough time for learners to be taught a whole language, let alone strategically reason out all of the possible patterns and forms in a language. Consequently, most of learning has to take place incidentally. This view has been advocated elsewhere (e.g., VanPatten & Williams, 2007) and will be regarded here as the default for naturalistic L2 acquisition. Note that this is not to deny that intentional learning is involved in language acquisition, nor is it to deny that it may be important. Even in the L1 intentional learning is important. For example, children are intentional learners when they enter school, even in their native language. Moreover, some innate bias for cultural learning presumably makes language acquisition intentional in some sense, i.e., children try to learn to communicate with their parents and social group (Tomasello, 2003). However, in the sense that intentional learning implies deliberate learning of linguistic knowledge that has been consciously defined by the learner as a priori important to learn, language acquisition is typically not intentional.

2.3. Incidental learning generally results in a knowledge base that is largely available to consciousness

In this section, I review the common findings of over 40 years of incidental (or implicit) learning research. The bulk of research on incidental learning has been conducted in cognitive psychology using non-linguistic stimuli, such as artificial grammars and serial reaction time tasks. As a consequence, much of the evidence—but not all the evidence—reviewed here will come from studies that often do not employ linguistic or even language-like stimuli. Nevertheless, the learning mechanisms and knowledge acquired in many of these tasks have been shown to elicit
similar behavioral phenomena and neural activity as natural language (e.g., artificial grammars, Petersson, Forkstam, & Ingvar, 2004; Williams, 2010; sequence learning, Christiansen, Conway, & Onnis, 2012; Misyak & Christiansen, 2012).

My aim here is to show that one of the most robust findings from the incidental learning literature is that conscious knowledge is often acquired under incidental learning conditions. This is not to negate the operation of implicit mechanisms on incidentally-acquired conscious knowledge. Indeed, as observed earlier in the discussion of PARSER and the SOC (Perruchet & Vinter, 2002), the development of lower-order explicit knowledge stems from the interplay between ongoing conscious experience and the operation of implicit mechanisms. Rather, in this review I simply mean to emphasize the conscious aspects of what is learned during incidental learning.

2.3.1. How is incidentally acquired knowledge represented?

2.3.1.1. Abstract rules

Most empirical research into incidental learning (i.e., learning scenarios in which participants are not forewarned of an impending test) has come from implicit learning research. In this paradigm, researchers typically are interested in whether incidental learning is possible and whether it results in a knowledge base that is inaccessible to awareness. The field of implicit learning began with Reber’s (1967) research in which he found that participants trained to memorize artificial grammar strings were able to classify new strings at levels better than what would be expected.

Note that incidental learning in this context is complex, since participants (a) were argued to have learned something apart from memorized fragments of training strings and since participants (b) have been shown subsequently to rely heavily on memorized fragments at test. However, other forms of information are undoubtedly at play over and above memory for fragments (e.g., repetition patterns, Pothos, 2007).
by chance. Moreover, participants were unable to verbalize the rules of the artificial grammar. Thus, Reber concluded that participants had internalized the constraints of the grammar: they had learned abstract, implicit rules. But does the evidence warrant this conclusion? Common everyday experience tells us that rules exist. Rules are why we recognize that “24683 is an odd number, and why Priscilla Presley is a grandmother…know that an offspring of raccoons that looks and acts like a skunk is nonetheless not a skunk” (Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995, p. 245). There is ample evidence that humans can learn and apply such rules, one only need consider scientific principles or mathematics. However, these rules are conscious. What evidence exists for the incidental learning of abstract, implicit rules?

Perhaps the most compelling evidence for rule learning has come from transfer studies, where participants are trained on stimuli from one modality and then classify stimuli in a different modality. In a classic study, Reber (1969) trained participants on artificial grammar letter strings in one modality by asking them to memorize them. Unbeknownst to participants, they were then switched to either a different grammar with the same letters (old lexicon, new rules) or different letter strings with the same grammar (new lexicon, old rules). Reber found that changing the rules disrupted memorization, while changing the lexicon did not. Reber interpreted this result as evidence that learners transferred implicit rule knowledge new letter strings. From this result, Reber concluded that incidental learning results in abstract, implicit rule knowledge. This transfer effect has been robustly replicated (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gomez & Schvaneveldt, 1994).

For further transfer-based evidence of abstract, implicit rule learning, Marcus et al. (1999) showed that seven-month-old infants were able to transfer knowledge of pseudoword

---

21 It is important to keep in mind that evidence for transfer does not necessitate a rule-based explanation, but only requires some level of abstraction for generalization.
sequences (e.g., ga-ti-ti and wo-fe-wo) following ABA and ABB patterns to novel pseudoword sequences that also followed the same patterns. Infants discriminated sequences with the same (grammatical) and different (ungrammatical) patterns even though they were unlikely to have conscious knowledge of the rules. Presumably such knowledge in seven-month-old infants can only be implicit.

Although these studies appear to provide evidence of abstract, implicit rule learning, they are limited in two ways. First, other studies have demonstrated that these same results do not necessitate an abstract, implicit rule-based explanation. For example, Redington and Chater (1996) demonstrated that memory for fragments of two or three letter combinations (i.e., bigrams and trigrams) was sufficient to explain participants’ abstraction behavior. If we assume that lower-order explicit knowledge can encode somewhat complex, abstract information, then we may not need to invoke abstract, unconscious rule computations. For instance, Marcus et al. (1999) showed transfer from one syllable set (e.g., gatitī) to another (e.g., wofeфе), where there is no natural correspondence between syllables, i.e., no reason to pair ga and wo or ti and fe. However, lower-order explicit knowledge can encode abstract, but simple, relations, such as repetition and change (e.g.,; Knowlton & Squire, 1996; Perruchet & Vinter, 2002; Williams, 2009, p. 330). Thus, the results do not require an interpretation exclusively in terms of implicit knowledge. A more pressing issue is whether rules are necessary to explain Marcus et al.’s results at all. Indeed, there have been multiple studies demonstrating that simple associative learning mechanisms can account for these findings, obviating the need for a rule-based explanation (e.g., Dominey & Ramus, 2000; Endress, Nespor, & Mehler, 2004; Pacton, Perruchet, Fayol, & Cleeremans, 2001). Finally, these results are largely unsurprising if participants developed higher-order explicit rule knowledge (which was unlikely in children, but
fully possible in adults in, for example, Reber’s studies). As noted earlier, the existence of conscious rules is not in question. It is fully possible to learn abstract, conscious rules. However, abstract rule knowledge as “linguistic competence” requires that these rules be implicit (e.g., Chomsky, 1965; VanPatten, 2012). Is there any evidence for the incidental learning of such abstract, implicit rule knowledge?

One convincing demonstration of abstract, implicit rule knowledge comes from Knowlton, Ramus, and Squire (1992) who compared artificial grammar learning in amnesic individuals with control learners. Amnesics and cognitively normal participants were trained on an artificial grammar by viewing letter strings and then writing them down. They were then given a recognition memory task. After this, participants were given a grammaticality judgment task. Amnesic patients were selectively impaired at making recognition judgments, but were able to perform as well as controls on a grammaticality judgment task, suggesting that they had an implicit knowledge base. That is, without apparently remembering specific training stimuli, amnesic participants could still correctly classify grammatical and ungrammatical items. Knowlton et al. (1992) suggested that this finding demonstrated the existence of an implicit knowledge base which was not tied to surface features of the stimuli, i.e., abstract, implicit rule knowledge. Similar findings with amnesic patients in Knowlton and Squire (1994, 1996) reinforced the conclusion that rules were unconscious, as they were clearly inaccessible for the amnesic participants. However, this interpretation of the results was called into question by Kinder and Shanks (2003), who trained and tested a simple recurrent network (SRN) on the same stimuli as participants in Knowlton et al. (1992). Kinder and Shanks made a theoretical assumption that amnesia was a general memory deficit leading to a slower rate of learning (rather than a selective memory deficit inhibiting only one subtype of memory). To simulate the
performance of amnesic patients, Kinder and Shanks simply lowered the learning rate of the SRN (cf. McClelland & Rumelhart, 1986). With this simple manipulation, Kinder and Shanks were able to replicate the amnesic and control results from Knowlton et al. (1992). Why does this cast doubt on the interpretation of amnesics learning implicit rules? First, the SRN cannot generally learn rules. It can only mimic rule-like behavior. Second, SRN consists of only a single learning mechanism, rather than having distinct rule-based and exemplar-based learning mechanisms. Yet, it is able to reproduce both rule-like and exemplar-like performance in humans. Thus, Kinder and Shanks argued that dissociations between apparently rule-based performance and exemplar-based performance follow from differences in learning rate in a single associative learning system. As such, the performance of amnesic patients is consistent with that of abstract, implicit rule knowledge but is also fully consistent with learning and memory systems that do not involve implicit rules. Therefore, the evidence provided by Knowlton et al. (1992) does not require a rule-based explanation.

Another widely-cited study interpreted as evidence for abstract, implicit rule knowledge comes from Meulemans and Van der Linden (1997). The authors reported that after different amounts of artificial grammar training, participants shifted from relying on exemplar specific information (e.g., chunks) to rule knowledge. However, the authors did not take care to unconfound rule knowledge from anchor chunk knowledge. Johnstone and Shanks (1999, 2001) demonstrated that participants could have performed well on the grammaticality judgment task just using fragments of string-initial letter combinations (e.g., all the legal strings began with MV, MX, VM, or VX). All that was required for attested level successful performance was conscious bigram knowledge.

---

22 Rule knowledge was operationalized as the contribution of the variable Grammaticality to classification performance on a grammaticality judgment task after all other variables had been accounted for.
Within SLA, Rebuschat (2008; Rebuschat & Williams 2012) investigated whether adult learners could learn L2 syntax under incidental learning conditions. They exposed learners to a semiartificial language consisting of English words placed into three German syntactic structures (e.g., *Yesterday bought John the newspaper in the supermarket*). After exposure to 120 training exemplars (40 of each structure), participants were given a surprise grammaticality judgment task (GJT) consisting of 60 new sentences, only half of which were grammatical. After each GJT trial, participants were asked to indicate their confidence in their decision (i.e., *guess, somewhat confident, or very confident*) and the source of their decision (i.e., *guess, intuition, memory, and rule knowledge*). These subjective measures (Dienes & Scott, 2005) were designed to assess the guessing criterion and zero correlation criterion of implicit knowledge. In short, if participants perform above chance when they indicate that they are guessing, then they have implicit knowledge. Likewise, if participants’ confidence is unrelated to their accuracy, then participants have implicit knowledge. The logic behind these measures is that participants who satisfy the guessing criterion and zero correlation criterion do not know that they are using the knowledge they have acquired, which implies that they do not know they have that knowledge. This judgment knowledge is implicit, as is the content of the knowledge that they are unaware of possessing, so-called structural knowledge (Dienes & Scott, 2005). Finally, when participants indicate that the basis of their decision is a guess or intuition, it is reasoned that participants’ structural knowledge is implicit. They do not have conscious access to the content of their knowledge. This is consistent with the phenomenology of a grammaticality judgment task. Ask a native speaker if *The dog ran down the street* is acceptable and s/he will say yes. Ask why it is grammatical and s/he will reply that it “just sounds right.” In essence, s/he is reporting intuition.
Rebuschat and Williams found little evidence that learners acquired abstract, implicit rules since there was no categorical performance, although some participants did appear to acquire microrules. Moreover, there was little evidence that participants acquired complete conscious rules, either. Some participants were able to verbalize some aspects of the training stimuli, but not enough to be considered abstract, conscious rules under the definitions outlined in Chapter 1. However, there was evidence that participants learned abstract, implicit knowledge; this knowledge did not appear to be rule-based, but instead may have taken the form of syntactic constructions (Rebuschat & Williams, 2009, 2012).

In a study on learning morphological form-meaning connections, Williams (2005) exposed learners to sentences containing artificial determiners under incidental learning conditions. Participants were informed that the four determiners (gi, ro, ul, and ne) indicated distance, but they were not informed of a hidden animacy regularity. After exposure to over 200 exemplars, participants were given a surprise forced-choice task, in which they had to choose the correct determiner. Crucially, in this task participants could only choose between determiners that had the same distance value, leaving animacy as the only deciding factor. Williams found that even participants who did not appear to be aware of the animacy regularity still performed above chance, suggesting that they had acquired some implicit knowledge of the animacy rules. It is important to keep two limitations to this conclusion in mind: first, participants who appeared to acquire implicit knowledge of the animacy rule did not perform categorically, which runs counter to the application of a rule. Second, participants were not given a generalization test, so their ability to apply their apparent rule knowledge was not assessed. Moreover, Williams also found that participants who had acquired higher-order explicit knowledge of the animacy regularity (i.e., they could report part or all of the animacy rule) performed well above
participants who had apparently acquired implicit knowledge. Replications of Williams’ (2005) study have found the same superior accuracy for participants with some degree of conscious knowledge over participants who appear to be relying on implicit knowledge (e.g., Hama & Leow, 2010; Rebuschat et al., in press). Moreover, neither replication study found any evidence of implicit rule learning, although Rebuschat et al. (in press) did find evidence for categorical application of abstract, conscious rules. Taken together, these studies do not show strong evidence of abstract, implicit rule learning under incidental conditions.

In sum, the classical assumption that incidental (or implicit) learning results in abstract, implicit rule knowledge has been undermined from a variety of directions (for overviews, see Perruchet, 2008; Pothos, 2007; Shanks, 2005). Importantly, the strongest criticisms have not been directed at the abstract nature of abstract, implicit rule knowledge. Rather, criticisms have primarily been directed at the tendency to equate abstract knowledge with rule knowledge (e.g., Perruchet & Poulin-Charronnat, forthcoming; Redington & Chater, 2002; see also lexicalist criticisms of rule-based syntax, e.g., Culicover & Jackendoff, 2006; Goldberg, 2006; Hagoort, 2005; Pollard & Sag, 1994; Tomasello, 2003). Infants and amnesic adults are able to incidentally acquire knowledge and transfer it to novel stimuli. Such abstraction, although consistent with a rule-based view, does not mandate a rule-based view, provided one bears in mind that non-rule-based knowledge can be abstract and that elementary learning mechanisms can be surprisingly powerful. That said, much of the evidence suggests that if participants do acquire rules, they are probably micro-rules (e.g., “the letter strings can begin with MV, MX, VM, or VX, but not VT.”). These rules are conscious, and vary in their level of abstractness.
2.3.1.2. Microrules

While much of the empirical evidence suggests that learners are unlikely to acquire knowledge of abstract implicit rules, this does not mean that participants acquire no rules. Rather, learners may acquire idiosyncratic microrules that often correlate with the descriptive rules used by the experimenter or by linguists. This notion was developed by Dulany and colleagues (Dulany, Carlson, and Dewey, 1984, 1985). They argued that during artificial grammar learning, participants not only acquire lower-order explicit knowledge of bigrams and trigrams (i.e., chunks), they also develop higher-order explicit knowledge, which is to say, they develop the ability to reflect back on the chunks they had acquired. In other words, learners may develop “personal sets of conscious rules, each of limited scope and many of imperfect validity” (Dulany et al., 1984, p. 541). They then use these so-called microrules in order to classify items in a test phase. It is even plausible that learners may only acquire knowledge of chunks during training, but when confronted with a surprise test, they then develop microrules from their chunks during the test phase. There is some evidence for this type of test phase-based development of microrules (e.g., Rebuschat, Hamrick, Sachs, Ziegler, & Riestenberg, in press).

Dulany et al.’s (1984) original study asked participants to indicate the basis of their grammaticality judgments by underlining parts of the artificial grammar string that made it grammatical. For example, if a participant judgment MTTV as grammatical, they might have underlined MT, which meant that MT led to endorsing MTTV as grammatical. Dulany et al. (1984, 1985) found that what participants underlined was able to predict mean accuracy on the grammaticality judgment task with no room for residuals. In other words, there appeared to be no effect of any other kind of knowledge, implicit or explicit, on performance. Reber (1993) criticized these studies, noting that participants could have underlined anything in a grammatical
string and it would have correlated with grammaticality, since a grammatical string must be grammatical in its entirety. While this is undoubtedly true, it does not obscure the fact that many other researchers have reported that participants typically are able to verbally report at least some of the knowledge they used to classify test items (e.g., Hama & Leow, 2010; Hamrick & Rebuschat, 2011, 2012, 2013; Mathews et al., 1989; Rebuschat, 2008, p. 119; Rebuschat et al., in press; Williams, 2005). This includes Reber himself: “Specific aspects of the letter strings were often cited as important in decision-making…Introspections after abound with references that have abstract rule-like qualities. Subjects refer to what can (and cannot) be…” (Reber & Allen, 1978, p. 202). Taken together, these findings suggest that participants are aware, at least to some degree, of the knowledge they use to classify stimuli.

In sum, learners may develop personal sets of idiosyncratic, partially-accurate, conscious rules, which are called microrules. At the least, the evidence shows that participants often believe to be using them during the test phase. Importantly, microrules are not the only knowledge that learners acquire, as Reber (1993) argues. However, it remains an open question to what extent learners can apply that other knowledge in as robust a way as they appear to apply their microrules.

2.3.1.3. Statistical computations

Statistical learning is increasingly seen as a viable alternative to abstract rule accounts in language acquisition research and cognitive psychology. These accounts, bolstered by evidence from behavioral experiments and computational modeling, suggest that language learning under incidental conditions rests on implicit mental processes that exploit the probabilistic patterns inherent in language (for overviews, see Rebuschat & Williams, 2012; Romberg & Saffran,
For instance, in the seminal study by Saffran, Aslin, and Newport (1996), 8-month-old infants listened to an unsegmented speech stream that contained no cues to word boundaries apart from dips in transitional probabilities at the edges of pseudowords. Transitional probability refers to the likelihood that a given element is followed by another element in a sequence. In English, this is easily demonstrated in the phrase pretty baby (Saffran, 2003). The transitional probabilities between syllables within words (e.g., *pre* → *tty* and *ba* → *by*) are higher than transitional probabilities between syllables across word boundaries (e.g., *tty* → *ba*). Using this logic, Saffran et al. (1996) hypothesized that if infants can track such statistical information, then it would lead them to segment words from the speech stream where the transitional probabilities were lowest. This is precisely what they found. Learning in this study was almost certainly incidental: infants were obviously not informed that there would be a test phase. Moreover, it is very unlikely that infants were deliberately seeking to exploit statistical information and even more unlikely that they were doing so in the way that, say, a statistician would (but see Gopnik, 2010, who argues that implicit processes in children are on par with that of statistical inferencing). To further examine the incidental nature of learning, Saffran, Newport, Aslin, Tunick, and Barrueco (1997) investigated whether older children and adults could also incidentally find words in an unsegmented speech stream, again with only statistical cues to word boundaries. Subjects were given a cover task of creating pictures on a computer. They were told that “they were participating in an experiment investigating the influence of auditory stimuli on creativity” (Saffran et al., 1997, p. 102). They were not forwarned of a test, nor were they even instructed to pay attention to the auditory stimuli. The original finding was replicated in two new age groups: adults and children were able to segment the speech stream above chance and with about the same accuracy.
There is further evidence that humans are sensitive to statistical information in language, beyond speech segmentation. Numerous studies have demonstrated that children and adults can learn on the basis of statistical cues under incidental conditions in a variety of other linguistic domains, e.g., word learning (e.g., Hamrick & Rebuschat, 2011, 2012, in press; Yu & Smith, 2007), learning form-meaning mappings (e.g., Lany & Saffran, 2010), phrase structure learning (e.g., Saffran, 2001), and syntax learning (Thompson & Newport, 2007). The robustness of these findings has led many to posit that humans possess a powerful statistical learning mechanism (e.g., Aslin et al., 1998; Saffran, 2003) capable of extracting a variety of knowledge types.

Common to most of these studies is a conceptualization of statistical learning as statistical computation, which entails more than sensitivity to just frequency. For example, Aslin, Saffran, and Newport (1998) pointed out that simple frequency counts of words and partial-words in their speech segmentations may have been a cue for extracting words from the speech stream in their original study (Saffran et al., 1996). To test whether learning was due to simple frequency effects or true sensitivity to more complex statistical information, like transitional probability, they redesigned their stimuli so that frequencies were completely balanced across words and non-words (thus making frequency an unreliable cue for word segmentation). This left only transitional probabilities as cues for infants. The original result from Saffran et al. (1996) was replicated. The results were interpreted as evidence that infants performed implicit statistical computations in order to segment words from the speech stream (but see Perruchet & Vinter, 1998 and chapter 3 of this dissertation for an alternative view explanation based on consciousness and memory).

---

23 Note that in Perruchet & Vinter’s (1998) account, learning is still statistical, but is based on chunk formation rather than statistical computation.
Within SLA, Williams (2010; Williams & Kuribara, 2008) has produced persuasive evidence that adults compute statistics during incidental learning. Williams and Kuribara (2008) compared the predictions of a Universal Grammar (UG) account and an Emergentist account of early incidental acquisition of L2 syntax in Japanese. In this study, English words were placed into Japanese syntactic structures both with and without scrambling and were affixed with Japanese morphemes (e.g., John-ga pizza-o ate; John-NOM pizza-ACC ate; John ate pizza). Participants were exposed to these “Japlish” sentences under the guise of a plausibility judgment task and were then given a surprise grammaticality judgment task on new sentences. Participants performed particularly well on canonical word order patterns in new sentences, suggesting some learning of abstract sequential patterns. However, learning was not as robust on scrambling structures. Finally, Williams and Kuribara then showed that an SRN trained on the syntactic category sequences from Japlish was able to closely match human behavior. However, no other computational models were tested for comparison, which makes it difficult to interpret what the SRN’s goodness of fit means.

Extending this work further, Williams (2010) investigated whether low performance on scrambling structures in Japlish could be accounted for by simply increasing the amount of exposure participants received. Participants were exposed to the same Japlish training materials as in Williams and Kuribara (2008), only this time participants were compared on the amount of training they received (194 exposures versus 388 exposures). This time, Williams (2010) found that increased exposure led to significant increases in performance for old scrambling structures but still without generalization to new sentences. Williams explained this finding in terms of a similarity-based approach. Since participants’ performance was based on the input, no amount of repeated exposure to the same training set would change the statistics inherent in that input.
Following this idea, Williams conducted a second experiment using nonsense syllable classes (e.g., si/se/sa/so and pi/pe/pa/po) as analogs for Japlish syntactic categories (e.g., Horse-ni farmer-ga hay-o gave became to-ni so-ga pa-o ku). Thus, any nonsense syllable beginning with ‘s’ corresponded to the subject of the sentence and so on. Williams found that endorsement rates for grammatical and ungrammatical nonsense Japlish analogs significantly correlated with those in the first experiment ($r = .83, p < .01$). When combined with the findings in Williams and Kuribara (2008), the finding that learning based on analog syntactic categories correlated with performance suggests that the learning mechanisms were operating on linguistic category information (as opposed to surface structure).

Then, Williams again simulated human learning behavior by training a SRN, this time on the nonsense syllable class (i.e., syntactic category analog) sequences from the training phase of the second incidental learning experiment. The results showed that the SRN was able to account for approximately 96% of the variance in the human data in the second experiment with the meaningless nonsense syllables, but only 40% and 66% of the data in the 194 and 388 exposure groups, respectively, in the first experiment (the reduction in fit was, presumably, due to the influence of other linguistic factors). The overall results suggest that adults are able to learn L2 syntax in a way consistent with the statistical computations of the SRN, suggesting that humans possess some functionally comparable statistical computation mechanism. On the other hand, such a mechanism in humans appears to be constrained in ways that the SRN was not, potentially leading to important differences in performance.

In sum, there is good evidence that incidental learning results in a knowledge base that makes learners sensitive to the statistical structure of the input. But to what extent do the mechanisms themselves operate consciously? And to what extent is the resulting knowledge
conscious? I address the learning process first. Inasmuch as statistical learning actually entails statistical computations, the learning process itself will certainly be implicit since learners do not consciously perform statistical calculations while learning. Indeed, some computational models of statistical learning operate under the assumption that learners perform unconscious computations like that of a statistician (e.g. Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). However, such views have been criticized for their lack of constraints and psychological plausibility (e.g., Altmann, 2010). More psychologically plausible approaches suggest that implicit statistical computations can be thought of as unconscious mental processes that track predictive dependencies or as learning probabilistic cue-outcome relationships (i.e., contingency learning). Connectionist networks like the SRN provide one way of simulating these sorts of computations (e.g., Misyak, Christiansen, & Tomblin, 2010; Shanks, 1995; Williams, 2009, 2010; Williams & Kuribara, 2008).

The notion that learners perform unconscious statistical computations of one sort or another during learning is gaining traction in cognitive science. Crucially, this notion brings together implicit learning and statistical learning. Perruchet and Pacton (2006) suggest that implicit learning and statistical learning may in reality simply be two theoretical approaches to a single phenomenon. Conway and Christiansen (2006) go as far as joining the two in name: *implicit statistical learning*. As noted above, this makes intuitive sense, since the types of probabilistic cues learners appear to become sensitive to in statistical learning studies are far too complex to be computed consciously by learners. Thus, it stands to reason that if statistical learning entails statistical computation, the learning process itself must be implicit. Indeed, no one claims that the sorts of complex probabilistic inferences postulated in statistical learning accounts of language acquisition are done consciously.
But the focus thus far has been on the learning process. What about the product? Is the knowledge that results from statistical learning implicit or explicit?\(^{24}\) This question has only been investigated recently and in a limited set of domains. For example, Hamrick and Rebuschat (2011, 2012, in press) combined subjective measures of awareness with the cross-situational learning paradigm (e.g., Yu & Smith, 2007). In these experiments, learners were asked to judge the animacy of objects while listening to “distracting nonsense” through headphones (they were, in fact, pseudowords). In each training trial, participants saw two objects and heard two pseudowords. Unbeknownst to the participants, the pseudowords and objects co-occurred probabilistically (e.g., a pseudoword like *houger* co-occurred with an elephant 50% of the time, a pear 33% of the time, and a glass 17% of the time). With the positions of the objects on the screen and the ordering of the pseudowords randomized, participants’ only cues to pseudoword-object mappings were co-occurrence probabilities. After exposure to a set of 57 trials, participants were given a surprise test where they have to match pseudowords with their object referents. Following Rebuschat (2008), subjective measures were used on each test trial to assess awareness. It was found that participants developed robust conscious knowledge, and even some implicit knowledge. Thus, there is some evidence that statistical learning results in implicit and conscious knowledge. Interestingly, participants’ implicit and conscious knowledge both appeared equally sensitive to the statistical patterns in the input, suggesting that inasmuch as learning was supported by some statistical learning mechanism, that mechanism appeared to exert an equal influence on the accuracy of both implicit and conscious forms of knowledge (Hamrick & Rebuschat, in press). In other words, the dissociation between implicit and

---

\(^{24}\) There is a related question, which I address in chapter 3: Do we need statistical computations to become sensitive to statistical information? Or, more directly, do we need statistical computations to learn when only statistical cues are present for learning?
conscious knowledge that is often reported in the literature was not found for statistical word learning.

In another study on incidental (or implicit) statistical learning and consciousness, Franco, Cleeremans, and Destrebecqz (2011) trained adults on a speech segmentation task like that of Saffran et al. (1996). Franco and colleagues found that adults were able to incidentally learn to segment words from a running speech stream which only contained statistical cues to word boundaries. Participants received either 10 or 20 minutes of exposure and were then given a surprise recognition test, asking them to identify “real” words from the speech stream. Using a variant of Jacoby’s (1991) process dissociation procedure,25 participants were asked to selectively use the knowledge they had acquired. It was reasoned that if participants were aware of having knowledge (i.e., if they are conscious that they know something), then they will be able to apply it or not apply it when asked. If, on the other hand, participants were not aware of having knowledge, they would be unable to exert such control. Using this procedure, Franco et al. found that statistical learning resulted in conscious knowledge, which participants could use intentionally.

In conclusion, although the literature on this issue is scant at the moment, the current evidence suggests that statistical learning under incidental conditions results in at least some conscious knowledge, and possibly implicit knowledge as well; however, the whole issue necessitates further investigation with triangulated measures of statistical learning and awareness. At any rate, the general pattern of incidental learning remains the same: incidental learning—which almost certainly involves the operation of unconscious processes at some level,

---

25 This procedure asks participants to use their knowledge to perform a test phase task, and then asks participants to deliberately not use their knowledge to perform the task. It is reasoned that if participants cannot prevent themselves from performing above chance or controls when asked to do so, then they are applying knowledge that they are not aware of having (i.e., it is implicit).
possibly even statistical computations—results in a knowledge base that is often available to consciousness.

2.3.1.4. Chunks

Within cognitive science, there is little debate that learning leads, at least, to the formation of chunks (Perruchet, 2008, p. 608). Furthermore, numerous studies have demonstrated that learners incidentally acquire small chunks\(^{26}\) during exposure to stimuli, whether in artificial grammar learning (e.g., Dienes, Broadbent, & Berry, 1991; Gomez & Schvaneveldt, 1994; Johnstone & Shanks, 1999; Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990; Reber & Allen, 1978), serial reaction time tasks (e.g., Perruchet, Bigand, & Benoit-Gonin, 1997), or language (e.g., Rebuschat, 2008; Robinson, 2005). For example, in a now classic study, Perruchet and Pacteau (1990) compared the performance of participants who were exposed to artificial grammar strings under standard incidental learning conditions with another group of participants who were only given bigrams (i.e., chunks) of training strings. They found that participants in both groups performed comparably. Thus, bigram knowledge may be all that is needed to account for learning. Moreover, they found that participants exposed only to bigrams were more likely to reject ungrammatical strings with illegal bigrams than ungrammatical strings with legal bigrams in the wrong places. However, bear in mind that this only demonstrates that when bigram information is the only source of knowledge, it will suffice for learning. However, when there are other sources of information available, learners appear to acquire chunk knowledge in addition to positional information (e.g., Dienes et al, 1991; Gomez & Schvaneveldt, 1994; Schuck et al., 2012). Moreover, Perruchet and Pacteau’s (1990) study does not demonstrate that chunks can combine to create larger chunks, which is another source of knowledge that learners

\(^{26}\) Although this is not to say that participants only acquire chunks.
appear to use (e.g., Servan-Schreiber & Anderson, 1990). In short, chunk knowledge suffices to explain much of learners’ performance, but there are multiple ways to get the same performance accuracy without invoking chunks. Therefore, one must be careful to actually demonstrate that chunks were involved in performance—not just sufficient for performance.

So how does incidental (statistical) learning result in chunk formation? There are, broadly speaking, two views. In the first view, the statistical computations discussed in the previous section result in chunks (i.e., sequences of items which have high co-occurrence probabilities with one another), which then can enter into new probabilistic computations with other chunks, and so on. This is the view taken by many in the statistical learning field (e.g., Fiser & Aslin, 2005; Saffran, 2001). To illustrate, imagine that the ngrams AB and CDE are units in an artificial grammar. Because A always predicts B, and C, D, and E mutually predict one another (the co-occurrence probabilities within chunks = 1.0), the implicit statistical learning mechanism will “chunk” them together as units. Thus, chunks emerge from statistical computations. Consider a similar example from the English sentence *Sheldon bought an air conditioner*. If statistical computations lead to the formation of chunks, then they should lead to the formation of the chunk *air conditioner* rather than *bought an air* because the co-occurrence probability between *air* and *conditioner* is very high, while the co-occurrence probabilities between *bought* and *an air* are intuitively lower. On this view, in natural language chunks may emerge as frequently co-occurring morphemes (e.g., *walked*) or words (e.g., *air conditioner*).

In the second view, incidental learning of chunks stems from the interplay between consciousness and memory, with the resulting chunks correlating to the statistical structure of the input. This idea is a part of the Fundamental Similarity Hypothesis (Robinson, 1996, 1997), which proposes that conscious and memory are involved in all aspects of L2 development. For
Robinson, incidental learning proceeds via the “accumulation of conscious processed, fragmentary, associative knowledge” (Robinson, 2013). This view is also consistent with Nick Ellis’ account of chunking in SLA (e.g., N. Ellis, 1998), in which chunks are formed by general purpose mechanisms like attention and working memory and are then operated on by implicit learning mechanisms.

These views are perhaps most clearly elucidated in the computational model PARSER, and the theory of self-organizing consciousness (e.g., Perruchet & Vinter, 1998, 2002). In PARSER, chunks are formed, initially, on a random basis by the limited capacity of attention, which constrains the contents of consciousness. In other words, the attentionally-constrained contents of consciousness form chunks. These chunks then gradually strengthen or weaken through general principles of associative memory. To illustrate, let us again imagine that we have an artificial grammar made up of bigrams. This time, the legal bigrams are AB, CD, EF, and GH. Initially, since PARSER does not know the structure of the input, it will extract random chunks of varying sizes from a series of processing episodes, with each processing episode consisting of conscious content (the corresponding experience in humans would be, say, conscious experience of ABC as a unit). Chances are that some of these chunks will be legal bigrams, while other will not be.

Consider a very simplified example. With no cues to segmentation, PARSER might extract the units AB, CDE, and FGH from the string ABCDEFGH. In this case, PARSER has created units in memory for one correct bigram (AB) and two illegal trigrams (CDE and FGH). These units in memory now guide the processing of the next sentence string: CDGHEFAB. Assume for the sake of simplicity that PARSER experiences this string as three more discrete chunks: CDG, HEF, and AB. In the processing of this second string, neither of the original
illegal trigrams CDE or FGH has been repeated in the input, so their memory traces decay. Moreover, the fact that these trigrams overlap with the recently-processed trigrams CDE and FGH means that the original illegal trigrams in memory are further forgotten due to interference. Finally, the legal bigram AB that was chunked in the first pass again matches the bigram AB in the second pass, leading to the strengthening of the AB bigram in PARSER’s memory. As this process repeats on a large corpus, PARSER is able to converge on the correct units of the input. Moreover, it has been shown that the structure of these units tends to correlate with the statistical structure of the input (Perruchet & Desaulty, 2008; Perruchet & Peereman, 2004). I return to this point in chapter 3.

Now, how do we distinguish these competing chunk formation processes in incidental learning? First, let us reconsider an example of incidental learning from earlier: Saffran et al.’s (1996) study of infant speech segmentation. Recall that in that study, infants listened to an unsegmented speech stream consisting of nonsense words (e.g., pabiku, golati), with no cues to word boundaries except dips in transitional probabilities. At test, infants were able to discriminate words from nonwords, suggesting that they had implicitly tracked statistical contingencies and extracted the correct word units. The first view of chunk formation that we mentioned certainly fits the findings: infants may have performed implicit statistical computations and then the outputs of those high-probability sequences were chunked by a bracketing mechanism (Giroux & Rey, 2009; Saffran, 2001). This is the type of process achieved in a connectionist SRN, which predicts what comes next in a sequence and, after learning, submits its outputs to a hierarchical cluster analysis, which forms chunks from high probability sequences (e.g., Elman, 1990). But are such computations necessary for chunk formation? In a series of studies, Perruchet and colleagues (e.g., Perruchet & Peereman, 2004; Perruchet &
Tillman, 2010; Perruchet, Tyler, Galland, & Peereman, 2004; Perruchet & Vinter, 1998) employed the PARSER model and accurately simulated a variety of incidental learning experiments, including those of Saffran et al. (1996, 1997), suggesting that statistical computations are not necessary to explain learners’ sensitivity to statistical structure in the input. In a more recent study, Giroux and Rey (2009) trained a SRN and PARSER on a speech segmentation task and compared the predictions of these models with human performance. They found that PARSER fit the human data better. This result suggests that PARSER may better account for some aspects of word segmentation than SRNs. More broadly, these findings suggest that chunk formation processes that are rooted in the dynamic interplay between conscious episodes and general principles of associative memory may suffice to account for incidental learning scenarios that have been previously thought to necessitate statistical computations (e.g., N. Ellis, 1998, 2005). This reliance on attested general principles of human cognition make PARSER appealing for its parsimony. No additional computational mechanisms are required beyond attention, awareness, and memory.

Finally, we must revisit the question of the conscious status of incidentally-acquired chunk knowledge. In chapter 1, I noted that chunks have both conscious and non-conscious content, consistent with PARSER and the self-organizing consciousness theory of Perruchet and Vinter (1998, 2002) and the signal detection theory model of Shanks and colleagues (e.g., Shanks & Berry, 2012). Importantly, there is no consensus regarding the conscious status of chunks within SLA. For example, traditional formulaic sequence accounts of chunks hold them to be conscious lexical entries (e.g., Wray & Perkins, 2000). On the other hand, usage-based approaches to SLA hold that chunks can be conscious, but also abstract and implicit (as in construction grammar, N. Ellis, 1998, 2005, 2006; Pedersen & Cadierno, 2004; see also non-
SLA accounts such as the Memory, Unification, and Control model, Hagoort, 2005). As such, it is important to review the evidence for these alternative accounts of chunks.

In a series of classic artificial grammar learning studies, Knowlton and Squire (1994, 1996; Knowlton et al., 1992) exposed amnesic patients and normal individuals (controls) to an artificial grammar, which was followed by an unexpected grammaticality judgment task and recognition memory task. Knowlton and Squire found that chunk strength influenced the grammaticality judgments of amnesic patients and controls to the same degree, i.e., chunk frequency predicted the participants’ classification decisions. Recognition memory for chunks was good in control participants, but was diminished greatly in amnesic patients. Now, since amnesia is an impairment only to a recognition mechanism, and not to a classification mechanism, it is feasible that both amnesic and control participants were able to use the same chunk knowledge to classify items in the grammaticality judgment task. On this view, classification is, by definition, implicit in amnesic participants, since they cannot be conscious of knowledge that they do not know they have. Thus, it appears that chunks can exert an influence on behavior without learners being aware of it, and this allows us to conceptualize chunks as a type of implicit knowledge.

On the other hand, chunks are not necessarily strictly implicit. A number of studies have shown that participants are conscious of chunks (e.g., Franco et al., 2011; Johnstone & Shanks, 1999; Mathews et al., 1989; Perruchet & Pacteau, 1990; Perruchet et al., 1997). In natural language, chunk knowledge is often lexically-specified and is therefore at least partly conscious. For example, people are fully aware of the lexical content of formulaic language (e.g., *in other words, my point is that X*), lexical bundles (e.g., *would you mind...*), and idiomatic constructions (e.g., *X kicked the bucket*). Moreover, such knowledge can be acquired incidentally in naturalistic
L2 settings (e.g., Webb, Newton, & Chang, 2013) and is processed as whole units in memory (e.g., Conklin & Schmitt, 2008). Moreover, several prominent theoretical accounts of natural language syntax now consider syntax itself to be represented as abstract chunks which are linked to lexical items (e.g., Culicover & Jackendoff, 2006; Goldberg, 2006; Hagoort, 2005), and there is empirical evidence to support this account (e.g., Kidd et al., 2010; Longe, Randall, Stamatakis, & Tyler, 2007; Novick et al., 2003; Reitter et al., 2011; Vosse & Kempen, 2000). Thus, there may be no contradiction in claiming that chunk knowledge could implicitly influence classification behavior and consciously influence recognition behavior. On this account, chunks are neither purely implicit nor purely conscious.

In sum, a wide range of incidental learning findings, from classic artificial grammar learning studies (e.g., Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990; Reber & Allen, 1978) to more recent work within statistical learning (e.g., Saffran et al., 1996, 1997) suggests that chunk formation is a common outcome of incidental learning, although it is unlikely to be the only outcome. Finally, regardless of the nature of the learning mechanisms involved, there is good evidence that the acquired chunks are consciously available to (cognitively normal) learners, since they typically can identify them in a recognition memory task (e.g., Perruchet et al., 1997) or can verbalize them to some degree (e.g., Dienes et al., 1991; Mathews et al., 1989; Reber & Allen, 1978).

2.3.2. Representation of knowledge acquired under incidental learning conditions

In this section, I have reviewed evidence regarding what is learned during incidental learning tasks and whether it is conscious. While there is very little evidence for incidental learning of implicit, abstract rules, there is a good deal of evidence for incidental learning of
conscious rules or microrules (e.g., Dienes et al., 1991; Dulany et al., 1984, 1985; Mathews et al., 1989; Reber & Allen, 1978; Rebuschat et al., in press; Hama & Leow, 2010). Likewise, there is some evidence for incidental learning on the basis of implicit statistical computations (e.g., Williams, 2010; Williams & Kuribara, 2008), with the resulting knowledge being implicit, but these findings are limited to just a few studies, all of which have shown more robust effects for the incidental statistical learning of conscious knowledge (e.g., Franco et al., 2011; Hamrick & Rebuschat, 2011, 2012, in press). Finally, there is a good deal of evidence that incidental learning results in conscious chunks, although this is not to deny that chunks can be abstract and implicit (e.g., Dienes et al., 1991; N. Ellis, 1998, 2005; Gomez & Schvaneveldt, 1994; Johnstone & Shanks, 1999; Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990; Reber & Allen, 1978). Thus, the bulk of evidence from studies conducted in cognitive science, psycholinguistics, and language acquisition research indicate that learning under incidental conditions often involves the acquisition of conscious knowledge of one sort or the other.

2.4. Conscious L2 knowledge is of intrinsic interest for SLA research

It has been argued that higher-order explicit knowledge (e.g., metalinguistic knowledge), by virtue of being a separate module of cognition, would be unable to contribute much to the development of L2 competence, which is primarily an implicit knowledge base (Krashen, 1981, 1985; Schwartz, 1999). This view has met with stiff opposition, with many arguing that some kind of interface may allow the development of implicit L2 knowledge on the direct or indirect basis of higher-order explicit knowledge (e.g., DeKeyser, 1998; N. Ellis, 2005; Schmidt, 1990, 1994, 1995, 2001). Issues of consciousness and L2 development are not just limited to SLA theory, either. As Schmidt (1995, p. 2) notes, issues of consciousness bear on foreign language
pedagogy. Traditional classroom instruction emphasizes the provision of higher-order explicit knowledge, while more recent approaches emphasize the development of tasks that are less intrusive than metalinguistic instruction and are more likely to increase learners’ lower-order explicit knowledge (i.e., tasks that promote noticing). In both cases, researchers and teachers are interested in whether conscious knowledge promotes L2 development, and this underscores nearly thirty years of pedagogical research. Put simply, conscious knowledge is of intrinsic interest for SLA researchers.

But why has this interest in consciousness been maintained? One possible reason is that practical concerns regarding L2 pedagogy necessarily guide instructed SLA research. A focus on how best to provide learners with conscious L2 knowledge falls out naturally from classroom research often focused on optimizing learner outcomes. Moreover, there are instructed SLA researchers and teachers who maintain that there is generally a positive relationship between conscious knowledge and L2 performance. There are also researchers who believe that the provision of conscious L2 knowledge can lead, directly or indirectly, to the sorts of implicit linguistic competence that many learners strive to attain (e.g., DeKeyser, 2003, 2007; N. Ellis, 2005; R. Ellis, 2009). For these researchers it is important to understand both how and why this is seemingly the case. In this section, I review some of the evidence for this view, acknowledging that it is by no means the definitive, or even correct, conclusion. Rather, it underscores the need for a clearer understanding of the relationship between conscious knowledge and L2 development. In this section, I review studies from three broad “categories” of SLA research on awareness: (1) incidental L2 learning studies, (2) instructed or intentional learning L2 studies, and (3) awareness studies using on-line measures.
2.4.1. Incidental and implicit L2 learning studies

Like research from cognitive psychology, research in SLA has shown that a common result of incidental learning is a knowledge base that is available to consciousness. Moreover, it appears that in SLA research participants who incidentally acquire conscious knowledge generally perform better on test measures than those who appear to acquire implicit knowledge. For example, in the studies by Rebuschat (2008; Rebuschat & Williams, 2012) reported above, participants were exposed to a semiartificial language consisting of English words placed into three German syntactic structures (e.g., *Yesterday bought John the newspaper in the supermarket*). Participants were then given a surprise GJT with subjective measures of awareness. Rebuschat and Williams found that learners acquired conscious judgment knowledge (i.e., they knew that they had learned something), but also some unconscious structural knowledge (i.e., they were above chance when indicating *intuition* as the basis of their GJT decisions). Crucially, accuracy on the GJT was significantly higher when participants used conscious judgment knowledge. Likewise, when participants indicated *rule knowledge* (a category meant to reflect higher-order explicit knowledge) as the basis of their decision, they performed better than when they indicated any other basis for their GJT decisions. In short, although participants appear to have acquired some implicit knowledge of non-native syntax, their accuracy when using conscious knowledge was higher. Thus, consistent with the research outlined in section 2.3, incidental learning was shown to produce some conscious knowledge, which learners were able to use to improve their accuracy. However, this interpretation should be treated with great caution, since these studies used untimed grammaticality judgment tasks (which may tap more conscious forms of knowledge, R. Ellis, 2005) and brief training periods, both of which may have biased the data in favor of finding robust conscious knowledge.
Similar results were obtained in the domain of incidental L2 word learning by Hamrick and Rebuschat (2011, 2012, in press). Participants were exposed to pseudowords and their referents in a cross-situational learning paradigm. In the incidental exposure phase, they were told to judge the animacy of objects while listening to “distracting nonsense” through headphones. Participants were then given a surprise four-choice picture matching task, and subjective measures were again used to assess the conscious status of the resulting knowledge. It was found that participants incidentally acquired both implicit knowledge and conscious knowledge. However, participants’ performance on items where they indicated using conscious knowledge (i.e., memory) was significantly better than on items where they indicated using implicit knowledge (i.e., guessing or using intuition).

Similar findings were obtained in the original study by Williams (2005) and subsequent replications and extensions (e.g., Hama & Leow, 2010; Rebuschat et al., in press). In these studies participants were exposed to sentences with artificial determiners that followed a distance rule and a hidden animacy rule. After exposure to over 200 exemplars, participants were given a surprise forced-choice task, in which they had to choose the correct determiner. In all experiments, participants who had acquired higher-order explicit knowledge of the animacy regularity (i.e., they could report part or all of the animacy rule) performed well above participants who did not.

In sum, although there are few studies that have reported on the development of conscious L2 knowledge under incidental learning conditions using such measures of awareness, the ones that have tend to show that conscious L2 knowledge results in generally more robust accuracy than implicit L2 knowledge. This finding is somewhat consistent with results from
studies with different training conditions, like instructional treatments and intentional learning, to which I turn next.

### 2.4.2. Instructed and intentional L2 learning

Researchers interested in optimizing L2 learning have a standing interest in how higher-order explicit knowledge can be used to promote learner development (e.g., by making it faster) and/or aid in the creation of implicit linguistic competence in the L2 (e.g., by stretching learning beyond what might be possible in purely naturalistic settings. SLA researchers interested in these issues have typically investigated them by examining the effectiveness of different pedagogical materials or by seeing how well learners perform when the actively seek out metalinguistic information in service of learning. In this section, I briefly review some of the findings from both of these domains. Many of these studies report some sort of positive relationship between conscious knowledge and L2 performance, albeit with limitations (e.g., DeKeyser, 2003; Norris & Ortega, 2000; Roehr, 2008; Sanz & Morgan-Short, 2005).

For example, Robinson (1996) exposed adults to exemplars of simple and complex grammatical rules in English. Two groups (incidental and implicit memory) received a more “implicit” treatment. The incidental group was instructed to read the exemplar sentences for meaning, while the implicit group was instructed to memorize the sentences. Two other groups received more “explicit” treatment: A rule-search group was instructed to find the rules governing the exemplars, while an instructed group was given the rules. After exposure to 40 training exemplars, participants were then given a grammaticality judgment task, followed by a questionnaire targeting participants’ abilities to describe the rules. Robinson found that the
instructed group, which received higher-order explicit knowledge, outperformed all other groups on grammatical and ungrammatical simple items and grammatical complex items.

Similar findings have been obtained elsewhere when researchers provide learners with higher-order explicit knowledge before practice (e.g., Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; N. Ellis, 1993). DeKeyser (1995), for example, found that participants who were presented with higher-order explicit knowledge (i.e., metalinguistic rules) did better on test items involving categorical rules, but not on probabilistic rules. This result should hardly be surprising, though, given that rules apply categorically. Depending on the actual probabilities involved, it is impossible to achieve 100% accuracy on probabilistic structures when applying categorical rules.

N. Ellis (1993) found that participants instructed on mutation rules in Welsh and who then practiced using their knowledge outperformed participants who did not receive instruction on a grammaticality judgment task. Rebuschat (2008, Experiment 6, pp. 137-155) found that participants instructed to consciously search for the word order rules governing his semiartificial language outperformed those who were incidentally exposed to the same materials. However, grammaticality judgment tasks are metalinguistic tasks, and so, again, performance advantages for learners who received explicit instruction in the above studies may be due to biases in the measurements used. Sanz and Morgan-Short (2005) point out that a problem with some of these studies is that the training or practice phases of these studies do not clearly have task-essentialness (Loschky & Bley-Vroman, 1993). In other words, it is unclear whether task demands forced learners to use the targeted forms in a meaning-focused way to successfully complete the tasks. For example, de Graaff (1997) found that the provision of explicit instruction accompanied by exposure and practice led to more robust performance than exposure and
practice without explicit instruction. However, the practice phases rarely included tasks that required form-meaning mappings, reducing task-essentialness.

In sum, SLA research using pedagogical interventions generally shows more positive learning effects associated with explicit instruction (i.e., provision of higher-order conscious knowledge), at least for some types of forms and on certain measures of learning. This is broadly the same conclusion reached by DeKeyser (2003) and Norris and Ortega (2000). However, these studies often lack measures or designs believed to be better at capturing implicit knowledge (e.g., delayed post-tests; Norris & Ortega, 2000; Sanz & Morgan-Short, 2004, 2005). As a result, it is unclear to what extent the benefits of explicit instruction over and above so-called implicit instruction are simply products of experimental design.

2.4.3. Awareness studies using on-line measures

In this section, I briefly review studies that have used on-line measures of awareness (i.e., concurrent think-alouds). These studies, too, are often interpreted as evidence in favor of a positive influence of conscious knowledge on L2 development. That is, these studies often report that higher levels of awareness (e.g., higher-order explicit knowledge, or awareness at the level of understanding, Leow, 1997; Schmidt, 2001) tend to correlate positively with learning. For example, Leow (1997) used think-aloud protocols to investigate the conscious contents of beginner Spanish students while they worked on a crossword puzzle task. Learners were instructed to report their thoughts aloud as they completed the task. Upon completing the crossword puzzle, participants were given a multiple-choice recognition task and a written production task. Leow found that participants who appeared to have some higher-order explicit
knowledge (as evidenced by metalinguistic comments) performed better than those learners who simply verbalized the forms (i.e., had lower-order explicit knowledge).

Similar findings have been obtained elsewhere (e.g., Leow, 2000, 2001; Rosa and Leow, 2004; Rosa and O’Neill, 1999). For example, Rosa and O’Neill (1999) and Rosa and Leow (2004) showed that learners who simply mentioned the target forms in their think-alouds did not perform as well as those learners who verbalized higher levels of awareness. Likewise, Sachs and Suh (2007) used think-alouds to investigated learner awareness during synchronous computer-mediated communication (CMC). Learners interacted with the researcher on the computer, completing a series of tasks, including one in which learners told a story either with or without textually-enhanced recasts and with or without thinking aloud. In general, they found that higher levels of awareness correlated with higher accuracy.

Further complicating matters, it may be that asking participants to describe the contents of their conscious experiences influences their conscious processes or their learning processes or both. Such reactivity has been a source of debate, since several studies suggest that concurrent think-alouds are reactive (e.g., Goo, 2010; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, in prep; Sachs & Polio, 2007; Sanz, Lin, Lado, Bowden, & Stafford, 2009), while others do not (e.g., Egi, 2004; Leow & Morgan-Short, 2004). For example, Sachs and Polio (2007) found that learners who verbalized the contents of their awareness during the learning phase performed worse overall than participants who did not do the concurrent think-alouds. Likewise, Sanz et al. (2009) conducted a computer-assisted language learning study on adult learners of Latin. In the post-test phase, the authors found that concurrent think-alouds inhibited reaction time in a post-test GJT, but did not influence accuracy on the GJT. Bowles and Leow (2005) also report that concurrent think-alouds resulted in slower reaction times. This result is, perhaps, unsurprising,
since it may naturally take longer to perform a given task when doing a secondary task (thinking aloud). Rebuschat et al. (in prep) also found negative reactivity of concurrent think-alouds in a replication of Williams (2005) and Hama and Leow’s (2010) artificial determiner studies. Consequently, concurrent measures of awareness may improve or hinder performance on a given measure of learning such that it is difficult to clearly see whether conscious L2 knowledge is beneficial, harmful, or merely epiphenomenal to development. As I will argue in chapters 3, 4, and 5, this potential for reactivity is another reason why researchers need to use multiple measures of awareness in order to triangulate the effects of conscious knowledge on development.

2.4.4. Limitations

In sum, there is evidence which suggests that when learners have conscious L2 knowledge, they perform better than those who do not have conscious L2 knowledge (e.g., either “implicitly-trained” learners or learners who show no evidence of awareness of the relevant regularity). However, this conclusion should be treated with caution for at least three reasons. First, there are some studies that do not report benefits for higher-order explicit knowledge (e.g., Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Sanz, 2004; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996). For example, Alanen (1995) found that the group with, in theory, the most conscious knowledge (“Rule & Enhance” group) had lower scores on the grammaticality judgment task than other groups in the experiment (although this appeared to be due to inaccurate application of the learned rules). More robust effects for “implicit training” over the provision of higher-order explicit knowledge have also been shown in brain imaging studies. Morgan-Short et al. (2012) found no advantage for the provision of higher-order explicit
knowledge over implicit training conditions. Moreover, they also found that metalinguistic instruction did not result in native-like brain activation patterns, while implicit training did. Now, the results of studies showing no advantages for higher-order explicit knowledge should also be treated with caution, and should not be taken to mean that learning took place without awareness in the implicit training groups. Many of these studies did not assess awareness, let alone include a measure of lower-order explicit knowledge (e.g., recognition memory), so it may be that they compared more explicit conditions with less explicit conditions. More research needs to be done to clarify this matter.

A second reason for treating these findings with caution is that the brief treatment/learning phases and test types used in many—but not all, see DeKeyser (1995)—of these studies favor the use of conscious knowledge (Norris & Ortega, 2000; Sanz & Morgan-Short, 2005). For example, treatment, learning, and exposure phases that last for mere minutes are widely regarded as being unable to produce implicit knowledge robust enough to support performance in the way that conscious knowledge does. This does not mean that participants have not acquired implicit knowledge, since they may have implicit knowledge which they cannot apply yet in a reliable way (Cleeremans, 2007). On the other hand, there are studies that do report implicit knowledge development after a relatively short exposure period. For example, Williams (2005) and Rebuschat (2008; Rebuschat & Williams, 2012) report that participants form implicit knowledge after between 120 and 250 exposures to exemplars, which, while fairly long for a laboratory study in implicit learning, is not much input by any real-world measure. Exposure to a few hundred exemplars simply may not be enough to extract the relevant implicit knowledge, let alone extract implicit knowledge that is robust enough to support high accuracy levels. The same is true for the above studies using pedagogical interventions and think-alouds.
However, it should be pointed out, especially in the latter case, that learners in think-aloud studies often do best when they have developed higher-order explicit knowledge (e.g., Leow, 1997; Sachs & Suh, 2007), even though they may also have some lower-order explicit knowledge. Thus, accurate, robust application of lower-order explicit knowledge, like implicit knowledge, may take longer to achieve, which skews the likelihood of more accurate performance in favor of higher-order explicit knowledge.

A third complicating issue is the complexity of the target forms. Some forms are simple, others are complex. Not all simple forms are simple in the same ways and not all complex forms are complex in the same ways. Some assume that simple forms are able to be learned explicitly (e.g., through instruction or intentional learning), while complex forms are better learned implicitly (e.g., DeKeyser, 1995). On the other hand, some claim that the opposite is true: implicit learning works better for simple forms, while complex forms require the development of conscious knowledge (e.g., Hulstijn & de Graaff, 1994). There is some empirical evidence for this as well. For example, Robinson (1996) found that explicit instruction produced more robust learning in the simple rules condition and the difficult rules condition. Moreover, there are some who point out that lower-order explicit knowledge is necessary for learning regardless of whether or not the form is simple or complex (e.g., Schmidt, 1990, 1994, 2001), and there is empirical evidence consistent with this view as well (e.g., Pacton & Perruchet, 2008; Spada & Tomita, 2010). For example, in a meta-analysis, Spada and Tomita (2010) found that effect sizes for explicit instruction were “consistently larger than those for implicit instruction,” on both simple and complex forms. Until we have a clearer picture of what kinds of forms are better learned through, say, explicit rule instruction versus simple exposure, we must keep in mind the
possibility that any benefit for conscious knowledge may be dependent upon the complexity of the linguistic targets.  

2.5. Summary

The primary aim of this chapter was to present an overview of the assumptions and observations underlying this dissertation. When considered together, these assumptions and observations lead to a gap in the literature. We know that most of naturalistic language learning is probably incidental, that incidental learning typically results in a knowledge base that is available, at least in part, to consciousness (i.e., often consisting of declarative memories for exemplars and higher-order explicit knowledge), and that conscious knowledge is of intrinsic interest to SLA researchers because of its complex relationship with L2 development. The gap comes from the fact that, despite having a vested interest in conscious knowledge and incidental learning, there is little SLA research aimed at investigating the incidental learning of conscious L2 knowledge in more detail.

Addressing this gap is interesting for psycholinguistic accounts of SLA. For instance, as pointed out in the last section, conscious knowledge is of intrinsic interest for SLA researchers, for both theoretical and practical reasons. Several accounts of SLA posit that conscious L2 knowledge, be it in the form of chunks or metalinguistic rules, can impact the acquisition of, apparently, implicit L2 competence, regardless of which interface position is taken (e.g., DeKeyser, 2003, 2007; N. Ellis, 2005; Ullman, 2004, 2005). Addressing this gap is also important for instructed SLA research. Traditional language pedagogy often involves the provision of higher-order explicit knowledge, while more recent approaches to language teaching

It is also crucial to consider that simple/complex and implicit/explicit dichotomies may oversimplify the reality of the matter, leading to difficulties in accurately interpreting research.
emphasize raising learner attention and awareness to form via more indirect ways, promoting lower-order explicit knowledge (e.g., from noticing) and learner autonomy in generating conscious knowledge. Thus, having a firmer understanding of the development of various types of conscious L2 knowledge stands to illuminate many aspects of SLA.

Finally, addressing this gap speaks to methodological concerns in SLA. Following calls for more triangulation in SLA research on awareness and learning (e.g., Leow, 2000; Robinson, Mackey, Gass, & Schmidt, 2012), this dissertation used three independent measures of conscious knowledge: recognition memory (e.g., Perruchet, Bigand & Benoit-Gonin, 1997; Shanks & Johnstone, 1999), retrospective verbal reports (e.g., Williams, 2005), and a novel subjective fluency judgment task. The results of the present experiments will shed light on the extent to which retrospective verbal reports capture conscious knowledge. If it is found that other measures of awareness show evidence of conscious knowledge beyond that found in the verbal reports, it will provide further support for the need for a more comprehensive battery of awareness measures. I return to this issue in Chapter 5.

2.6. The present experiments

The aim of the present experiments was to investigate a set of related questions regarding the development of conscious knowledge during early incidental learning of L2 syntax. The aim of the first experiment was to investigate the representation of the knowledge acquired during early incidental learning of L2 syntax (e.g., abstract rules, microrules, statistics, and/or chunks). This experiment did not directly address whether learners develop conscious knowledge, but, rather, assessed whether learners use learning mechanisms and acquire types of knowledge that are compatible with conscious representation (e.g., chunks, rules, microrules) or not compatible with
conscious representation (e.g., statistical computation). Building on the first experiment, the second experiment investigated the development of conscious knowledge under incidental learning conditions using three different measures of awareness designed to capture both lower-order and higher-order explicit knowledge: a recognition memory test, a retrospective verbal report interview, and on-line subjective fluency ratings (these are explained in detail in Chapter 4). This experiment provides insights on the development of conscious L2 knowledge under incidental learning conditions, including (1) at what point learners develop conscious knowledge, (2) whether they continue to passively memorize conscious knowledge with increased exposure (e.g., consistent with Robinson, 1996, 1997), and whether both lower-order and higher-order explicit knowledge are acquired.
3.1. Introduction

The aim of Experiment 1 was to elucidate the mechanisms supporting L2 syntactic development under incidental learning conditions, as well as the representation of the acquired syntactic knowledge. Since there is a consensus within the incidental learning and implicit learning research traditions that learning essentially involves the extraction of information from statistical regularities, the present experiment focuses on statistical learning of L2 syntax. The following research questions served as the basis of the experiment:

1. Can statistical learning of L2 (i.e., semiartificial) syntax be obtained under incidental learning conditions?

   Hypothesis 1: There will be evidence of statistical learning as indicated by speedup in reaction time measures and by accuracy in a grammaticality judgment task, consistent with previous research on statistical learning of syntax (e.g., Thompson & Newport, 2007) and incidental learning of L2 syntax (e.g., Rebuschat, 2008; Rebuschat & Williams, 2012; Williams, 2010; Williams & Kuribara, 2008).

2. Is there any behavioral evidence that adult learners perform statistical computations on their input?

   Hypothesis 2: There will be evidence of statistical computations during the exposure phase as indicated by speedup in reaction times for high-probability sequences compared with low-probabilities sequences, consistent with previous research demonstrating that reaction time gains that correspond to computing statistics (e.g., Hunt & Aslin, 2001).
3. Is the resulting knowledge consistent with statistical computation views of statistical learning (e.g., Aslin et al., 1998; Griffiths et al., 2010; Misyak, Christiansen, & Tomblin, 2010), with chunk formation views of statistical learning (e.g., Perruchet & Peereman, 2004; Perruchet & Vinter, 1998, 2002), neither or both?

Hypothesis 3: If hypothesis 2 is correct, then human performance will correspond with that of the SRN, consistent with previous research on statistical learning of L2 syntax (e.g., Williams, 2010; Williams & Kuribara, 2008).

Hypothesis 4: If hypothesis 2 is not correct, then human performance will correspond with that of PARSER, consistent with chunk-based accounts of language acquisition (e.g., N. Ellis, 1998) and previous research that has demonstrated superior performance for PARSER over the SRN in accounting for human behavior on psycholinguistic measures (e.g., Giroux & Rey, 2009; Perruchet & Peereman, 2004).

Before reviewing Experiment 1 in detail, I first briefly review issues in the literature that motivated the study.

3.2. Mechanisms of statistical learning

Both children and adults are sensitive to a range of statistical structures in language (for overviews, see Aslin & Newport, 2008; Rebuschat & Williams, 2012; Romberg & Saffran, 2010; Saffran, 2003). The seminal demonstration of human sensitivities to statistical structure comes from Saffran et al. (1996), who exposed 8-month-olds to unsegmented auditory strings consisting of pseudowords (e.g., golati, pabiku, henceforth “words”). Crucially, the auditory strings contained no prosodic or acoustic cues to word boundaries. The only cue to word boundaries was the strong statistical relationships between syllables within words relative to the weak statistical
relationships between syllables at word boundaries. Saffran and colleagues found that infants were able to discriminate “words” from part-words (e.g., *latipa*), indicating that the infants had used statistical information to segment the speech stream. Numerous studies have replicated these results and extended them to other domains of language learning, providing increasing evidence that infants (e.g., Aslin et al., 1998; Lany & Saffran, 2010) and adults (e.g., N. Ellis & Larsen-Freeman, 2009; Franco et al., 2011; Hamrick & Rebuschat, 2012, 2013; Thompson & Newport, 2007; Williams, 2010; Williams & Kuribara, 2008) are sensitive to statistical information.

While there is broad agreement that humans are sensitive to statistics, there is less agreement over the cognitive mechanisms that underlie these sensitivities. Some have argued that human sensitivity to statistics is the result of learning mechanisms which compute statistics over the input (e.g., Aslin et al., 1998; N. Ellis, 2005, 2006). Some researchers have taken this view so far as to posit that statistical learning is the result of learners performing unconscious computations like that of a statistician (e.g. Griffiths et al., 2010). However, such views have been criticized for their lack of constraints and psychological plausibility (e.g., Altmann, 2010). More psychologically plausible approaches suggest that implicit statistical computations can be thought of as unconscious mental processes that track predictive dependencies or as learning probabilistic cue-outcome relationships (i.e., contingency learning). Connectionist networks like the SRN provide one way of simulating these sorts of computations (e.g., Misyak, Christiansen, & Tomblin, 2010; Shanks, 1995; Williams, 2009, 2010; Williams & Kuribara, 2008). However, some argue that there are other ways to account for human sensitivity to statistics without invoking learning mechanisms that actually compute statistics. For example, the PARSER model (Perruchet & Vinter, 1998, 2002) predicts that human sensitivity to statistical structure is a
natural consequence of general principles governing attention, awareness, and memory. This leads to a radically different conception of statistical learning. On this view, statistical sensitivity is a byproduct of chunk formation processes and no dedicated statistical computation mechanism is required.

Importantly, these different models sometimes offer different predictions about linguistic phenomena, and thus provide a unique method for gaining insights into the mechanisms of language learning. Up to now, the bulk of research comparing these models has been carried out in the domain of word segmentation (e.g., Giroux & Rey, 2009; Perruchet & Peereman, 2004). For example, Giroux and Rey (2009) trained a SRN and PARSER on a speech segmentation task and compared the predictions of these models with human performance. They found that PARSER fit the human data better than the SRN, suggesting that PARSER may better account for some aspects of word segmentation than SRNs. Moreover, this finding is consistent with other studies showing that other chunk-based models outperform other statistical sequence learning models in other domains, such as visual pattern learning (e.g., Orbán, Fiser, Aslin, & Lengyel, 2008) and artificial grammar learning (e.g., Boucher & Dienes, 2003). Importantly, this suggests that advantages for chunk-based models over sequence learning models are not limited to the idiosyncrasies of a single model (i.e., PARSER) in a single domain (i.e., speech segmentation), but instead may be more general property of chunk formation itself as a method for learning.

One of the central aims of Experiment 1 was to assess how these different models account for L2 syntactic development under incidental conditions. Some modeling research has already been done on incidental learning of L2 syntax. For example, Williams (2010; Williams & Kuribara, 2008) was able to use SRNs trained on syntactic category analogues to account for
as much as 96% of the variance in adult learning of L2 syntax under incidental conditions, suggesting the presence of comparable mechanisms in adult learners. However, Williams only used the SRN to fit the human data, constituting a proof that the human behavior was, in principle, possibly the result of simple associative mechanisms computing predictive statistics. Williams did not compare human performance with any other computational model of statistical learning, leaving open the possibility that another model could better account for the data.

Experiment 1 examined precisely this possibility by comparing the performance of human adults with that of the SRN and PARSER in a study on L2 syntactic development under incidental conditions. However, before detailing the methods of the present study, I briefly review the key theoretical and architectural differences between the SRN and PARSER models that are compared in this chapter.

### 3.2.1. SRN

According the SRN model (Elman, 1990) learning takes place via predictive statistical computation. The SRN (depicted in Figure 3.1 below) is given a sequential input one item at a time. At each time step, the SRN makes a prediction about what comes next in the sequence. When the next item is input, the SRN compares that input with its previous prediction. If its prediction is correct, then the SRN adjusts its connection weights to increase the likelihood that it makes a correct prediction in the future. If its prediction was incorrect, though, then the SRN adjusts its connection weights to increase the likelihood that it does not make the same mistake again in the future.
Figure 3.1. Architecture of the simple recurrent network. The feedback loop between the hidden units and context units provide the network with a kind of short-term memory analog, which allows the network to predict what occurs next relative to what it is currently processing.

(https://plato.stanford.edu/entries/consciousness-temporal/Fig23.png)

To illustrate, consider giving the SRN an unsegmented stream of letters from an artificial grammar composed of the bigrams AB, CD, and EF. Now, let us consider a random concatenation of these bigrams: ABCDEFCDEFABEFCDAB. Descriptively, this string consists of bigrams whose letter sequences have transitional probabilities of 1.0 (e.g., B|A = 1.0, D|C = 1.0, and F|E = 1.0). However, the transitional probabilities between bigrams are considerably lower, e.g., CD|AB = 0.33, EF|AB = .33, etc. When given each letter, the SRN predicts which letter comes next. Ultimately, the SRN learns by adjusting its connection weights to maximally reflect the transitional probabilities in the input (e.g., N. Ellis & Larsen-Freeman, 2009; Perruchet & Peereman, 2004). In essence, the SRN learns to represent the statistics of the input
in its distributed set of connections. Because the probabilities within bigrams are higher than those between bigrams, the SRN will be able to learn the appropriate units of the artificial grammar.

3.2.2. PARSER

PARSER forms chunks on the basis of simulated attentional and awareness mechanisms. Chunks are formed initially on a random basis by the limited capacity of attention, which constrains the contents of consciousness. That is, the content of simulated conscious experience (i.e., what is noticed) forms a chunk that enters into PARSER’s memory. These chunks then gradually strengthen or weaken through general principles of associative memory. To illustrate, let us also apply PARSER to an unsegmented stream of letters from an artificial grammar, this time composed of the bigrams AB, CD, EF, and GH. Initially, since PARSER does not know the structure of the input, it will extract random chunks of varying sizes from a series of attentional processing episodes, with each processing episode consisting of “conscious” content (the corresponding experience in humans would be, say, conscious experience of ABC as a unit). Chances are that some of these chunks will be the legal bigrams, while others will not be.

Consider a very simplified example. With no cues to segmentation, PARSER might extract the units AB, CDE, and FGH from the string ABCDEFGH. In this case, PARSER has created units in memory for one correct bigram (AB) and two illegal trigrams (CDE and FGH). These units in memory now guide the processing of the next sentence string: CDGHEFAB. Assume for the sake of simplicity that PARSER experiences this string as three more discrete chunks: CDG, HEF, and AB. In the processing of this second string, neither of the original illegal trigrams CDE or FGH has been repeated in the input, so their memory traces decay. Moreover, the fact that these
trigrams overlap with the recently-processed trigrams CDE and FGH means that the original illegal trigrams in memory are further forgotten due to interference. Finally, the legal bigram AB that was chunked in the first pass again matches the bigram AB in the second pass, leading to the strengthening of the AB bigram in PARSER’s memory. As this process repeats on a large corpus, PARSER is able to converge on the correct units of the input.

3.3. Experiment 1

The present study had three primary goals. The first was to determine whether evidence of statistical learning of L2 syntax could be obtained under incidental learning conditions. The second goal was to determine whether the learning effect (assuming there was one) was due to statistical computations. The final goal was to compare two prominent models of statistical learning, SRN and PARSER, in order to see which could better account for the patterns of human performance.

In order to address these goals, a semiartificial language learning experiment was conducted on human adults. The experiment contained three innovations in the semiartificial language learning paradigm: First, it used a trained control group in order to avoid problematic assumptions involved in comparing experimental groups with untrained controls or chance baselines (e.g., Hamrick, 2012, 2013; Hamrick & Sachs, in prep; Perruchet & Reber, 2003; Rebuschat et al., 2013, forthcoming). The second methodological innovation concerned the development of the semiartificial language stimuli. In previous studies using semiartificial languages, the statistical structure of the stimuli was not controlled. So, even if statistical

---

28 The semiartificial language paradigm was used for two reasons. The first was comparability with previous studies looking at the mechanisms of L2 syntactic development. The second reason was a practical choice: using semiartificial languages reduces the duration of the experiment, since the words are usually in participants’ native language, they need not be pre-trained on vocabulary.
learning took place, it was unclear what types of statistics were informative for learners. Therefore, the present study employed a semiartificial language whose syntactic structure was constructed to have only two transitional probabilities between syntactic categories (67% and 33%). The third innovation was that results of the human experiment were compared with two computational models of statistical learning, rather than one: the SRN, which learns via predictive computation, and PARSER, which learns via chunk formation.

3.4. Method

3.4.1. Participants

Thirty volunteer undergraduate native speakers of English (21 women, 9 men, $M_{age} = 18.76$, range = 18-20) were randomly assigned to either experimental ($n = 15$) or control ($n = 15$) conditions. Data from four participants were discarded because they failed to follow task instructions ($n = 3$) or had prior knowledge of a language (Persian) whose syntax matches the structures used in the semiartificial language ($n = 1$), leaving thirteen participants in each group. Experimental and control groups did not differ significantly across age, sex, handedness, or number of languages (all $ps > .05$). All participants reported having normal or corrected-to-normal vision.

3.4.2. Stimuli

This section describes the materials for Experiment 1. There were two sets of exposure phase stimuli: an experimental set and a trained control set. Finally, there was a set of test phase stimuli that all participants read during the grammaticality judgment task.
3.4.2.1. Experimental stimuli

The experimental group was exposed to a semiartificial language consisting of English words and syntactic structures\textsuperscript{29} based on Persian. Three syntactic structures were used to generate 96 sentences (32 sentences per structure). To create the stimuli, simple transitive English sentences were rearranged according to the three syntactic structures while still obeying within-phrase structure rules of English\textsuperscript{30}, as in Table 3.1 below.

Table 3.1
Three structures used in the exposure phase of the experimental group and sentences exemplifying each. Note: the use of the same lexical items for each is only for illustration.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Syntactic Category Sequence</th>
<th>Sample Sentence</th>
</tr>
</thead>
</table>

Stimuli were balanced as carefully as possible within the confines of natural language. There were five repeating temporal phrases, twenty repeating subject proper nouns (all names), non-repeating prepositional phrases, non-repeating object nouns, and twice repeating verb phrases. These can be found in the appendix. Sentences were presented in random order in the training phase.

\textsuperscript{29} Syntactic phrase/category is used here only to denote a lexical constituent or constituents that constitute a single syntactic phrase in English. It may be that participants process argument roles or some more basic category structures rather than syntactic categories, per se.

\textsuperscript{30} This may lead to an alignment differences for combining the syntactic information in the lexicon of one language with the syntactic information from the syntax of the other. Whether or not this manifests as a problem is an empirical question. Thanks to Rebecca Sachs for bringing this to my attention.
The experimental sentences contained syntactic phrase sequences that were probabilistically constrained. The transitional probabilities between syntactic phrases were either .67 or .33, depending on the transition (see Table 3.2). Thus, each syntactic phrase had a distributionally preferred, but not mandatory, successor in each sentence (e.g., TEMPORAL PHRASE could be followed by SUBJECT, or VERB PHRASE, but was more likely to be followed by the former). Transitional probabilities were averaged over each experimental structure to calculate a mean transitional probability for each sentence type as a whole. These were as follows: Structure A = .58, Structure B = .42, Structure C = .50.\footnote{Incremental transitional probabilities were also tracked (e.g., the probability of $T \rightarrow S$ was .67, the probability of $TS \rightarrow P$ was .5, the probability of $TSP \rightarrow O$ was 1.0, and so on), but statistical analyses revealed no significant effect of incremental transitional probability on reading times (all $p$s > .10).}

### Table 3.2

<table>
<thead>
<tr>
<th>Transition</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMPORAL PHRASE – SUBJECT</td>
<td>0.67</td>
</tr>
<tr>
<td>TEMPORAL PHRASE – VERB PHRASE</td>
<td>0.33</td>
</tr>
<tr>
<td>SUBJECT – PREPOSITIONAL PHRASE</td>
<td>0.67</td>
</tr>
<tr>
<td>SUBJECT – OBJECT</td>
<td>0.33</td>
</tr>
<tr>
<td>PREPOSITIONAL PHRASE – OBJECT</td>
<td>0.67</td>
</tr>
<tr>
<td>PREPOSITIONAL PHRASE – VERB PHRASE</td>
<td>0.33</td>
</tr>
<tr>
<td>OBJECT – VERB PHRASE</td>
<td>0.33</td>
</tr>
<tr>
<td>OBJECT – PREPOSITIONAL PHRASE</td>
<td>0.33</td>
</tr>
<tr>
<td>VERB PHRASE – SUBJECT</td>
<td>0.33</td>
</tr>
</tbody>
</table>

#### 3.4.2.2. Trained control stimuli

Part of the novelty of the present study is the fact that the control group also participated in a training condition (for the importance of using trained control groups, see Perruchet & Reber, 2003). The control group was trained and tested on the same stimulus sentences as far as the lexical items and the compositional semantics contained in them were concerned. However, the
exposure phase sentences in the control group did not follow the three syntactic structures, but were randomized in such a way that no whole sentence structure was ever repeated (e.g., TSPOV only occurred once and temporal phrases only were sentence-initial for 1/5 of the stimuli). That is, each of the 96 sentences was presented in a different syntactic phrase order with transitional probabilities between syntactic phrases matched for all sequences. This manipulation meant that the transitional probability between any two syntactic phrases over the course of the training phase was approximately 0.25. Thus, there were no probabilistic cues to word order in the control stimuli. This provided a learning baseline that ideally isolated unforeseen task effects or mere exposure effects from the exposure phase, thus allowing them to be partialed out.

3.4.2.3. Test materials

The 36 novel test phase sentences consisted of the three target grammatical structures (A, B, and C from Table 3.1) and three ungrammatical structures (D, E, and F, from Table 3.3) all with new lexical items compared to the training phase. All test sentences were constructed with two temporal phrases, six subject proper nouns (names), six prepositional phrases, six object nouns, and six verbs. Thus, the same core set of lexical items was rotated around the test stimuli. This was done to limit over-acceptance or rejection of any one structure at test due to its lexical content, since each structure was drawn from the same set of lexical items. As with the exposure phase, test items were presented in random order across participants.
Table 3.3
Ungrammatical patterns used in the test phase. Sentences are taken from the same lexical items for illustration purposes only.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Syntactic Category Sequence</th>
<th>Sample Sentence</th>
</tr>
</thead>
</table>

3.4.3. Procedure

Each group first participated in an exposure phase (which differed only in terms of their stimuli) and completed identical test phases.

3.4.3.1. Experimental exposure phase

The exposure phase was set up as a plausibility judgment task within a non-cumulative self-paced reading design. Implausible sentences were defined for participants as those unlikely to happen in the real world (e.g. Yesterday John at the store milk bought is plausible, but Yesterday John at the store milk sang is not). Sentences in the exposure phase were divided into four blocks of 24 sentences each. Of those 24 sentences, each structure occurred eight times, half in semantically plausible sentences and half in semantically implausible sentences. All sentences were semantically plausible until the final word, which either maintained or violated plausibility. This ensured that participants would read the entire sentences. If implausible content was delivered too early in the sentence, participants would not have to process the entire sentence to complete the task, thus making learning less likely. The order of the four training blocks was randomized across participants, as was the order of the sentences within those blocks.
3.4.3.2. Trained control exposure phase

As with the experimental group, the control group was exposed to sentences in four randomized blocks. Likewise, half of each block consisted of semantically plausible sentences, while the other half was implausible. As stated previously, the crucial difference was that no sentence structure ever repeated in the control phase. Thus, the control group simply received four randomized blocks of randomly ordered, non-repeating sentence structures.

3.4.4. General procedure

Participants were tested individually in a quiet laboratory. They were told that they were participating in a study about meaning comprehension under time pressure (see appendix for full instructions). Participants were instructed to read through sentence fragments one at a time by pressing the space bar to reveal each new fragment, repeat the sentence aloud (to ensure they were paying attention), and then indicate whether the sentence depicted a scenario that was likely in the real world (i.e., make a plausibility judgment).

The experiment was administered on a PC with a 15.6” screen using SuperLab Pro 4.0 for Windows. All text was black on a white background with size 18 Tahoma font. Sentences were presented in a non-cumulative moving-window self-paced reading design (Figure 3.1), which required participants to press the space bar to present each fragment of a sentence. Importantly, the boundaries between fragments corresponded exactly with the syntactic phrase boundaries that had been probabilistically manipulated (that is, temporal phrase, subject, object, prepositional phrase, and verb). In other words, participants saw the constituents of a single complete syntactic phrase each time they pressed the space bar. This manipulation was intended
to prevent learners from having to perform extra computations to segment the primitive units of the sentence (syntactic phrases), meaning that transitional probabilities between syntactic phrases would no longer have been the only relevant probabilities for learning.

For each trial in the exposure phase, participants saw a fixation cross and then pressed the space bar to begin reading sentence fragments. They continued pressing the space bar to see each fragment until they reached the end of the sentence. They were then prompted to repeat the sentence verbatim. After doing so, they pressed the space bar again and were prompted for a plausibility judgment. After giving a plausibility judgment, participants saw another fixation cross and the next trial would begin. The average time taken to complete the exposure phase was 20 minutes.

<table>
<thead>
<tr>
<th>Event</th>
<th>Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Bar Press</td>
<td>Yesterday ---------------</td>
</tr>
<tr>
<td>Space Bar Press</td>
<td>-------------- Charlie</td>
</tr>
<tr>
<td>Space Bar Press</td>
<td>-------------- at the store</td>
</tr>
<tr>
<td>Space Bar Press</td>
<td>-------------- milk</td>
</tr>
<tr>
<td>Space Bar Press</td>
<td>-------------- bought</td>
</tr>
</tbody>
</table>

Figure 3.1 Example of the non-cumulative moving-window self-paced reading design.

After the exposure phase, participants were then told that the word order in the previous sentences was not random but instead had contained systematic patterns. They were instructed to read 36 new sentences, all of which would be plausible. They were told that half the new sentences would follow the same word order patterns as the previous sentences and that these should be called “grammatical.” They were told that the other half of the new sentences would not conform to the same word order and should be rejected as “ungrammatical.” To decrease the likelihood that they would classify sentences on the basis of their meanings, participants were
reminded that all test sentences were plausible and the focus now was on word order. On average, the test phase took 5 minutes to complete.

3.5. Results

Results are organized in terms of the research questions outlined at the beginning of this chapter. Alpha levels were set to 0.05 for all statistical analyses and post-hoc comparisons were only conducted when omnibus analyses were significant.

3.5.1. Research question #1

The first research question asked whether evidence for statistical learning of L2 (i.e., semiartificial) syntax could be obtained under incidental learning conditions. To answer this question, first reading times were analyzed, followed by the results of the grammaticality judgment task. These results are reported in turn below.

Hypothesis 1 predicted that there would be evidence for statistical learning of L2 syntax in the form a reading time speedup for the Experimental group during the exposure phase. To investigate this possibility, reading times for sentences were averaged for each participant within each block. A 2x4 mixed ANOVA with Block (four levels: Block1, Block 2, Block 3, and Block 4) as within-subjects variable and Condition (two levels: experimental v. control) as between-subjects variable revealed a main effect of Block (Greenhouse-Geisser corrected values), $F(1.6, 35.28) = 7.311, p < .01, \eta_p^2 = .25$. There was no effect of Condition, $F(1, 22) = .833, p = .37$, nor was there a significant interaction between Block and Condition, $F(1, 22) = .746, p = .45$. Thus, the results indicate that both Experimental and Control participants had significant decreases in reading time during the exposure phase (Figure 3.2). However, there is no evidence that the
Experimental group used the reliable probabilities in their input to become faster than would be expected from simple practice effects shown in the control group. Thus, there was no evidence for statistical learning of L2 syntax by overall reaction time measures.

Figure 3.2. Mean reading times across the four training blocks in the Experimental and Control groups.

The following analyses were conducted to see if there was any evidence of statistical learning in the Experimental group on the grammaticality judgment task. The analysis of the grammaticality judgments showed that the Experimental group classified 61.89% ($SD = 16.51$) of the test items correctly and the Control group 59.93% ($SD = 12.51$). The difference between the two groups was not significant, $t(24) = .35$, $p = .72$. In order to establish whether the Experimental and Control groups performed differently across the individual syntactic structures.
in the test phase, a 2x6 mixed ANOVA was conducted on participants’ accuracy on the grammaticality judgment task with Group (2 levels: Experimental, Control) as between-subjects factor and Structure (6 levels: A, B, C, D, E, F) as within-subjects factor (see Figure 3.3 below). The ANOVA (with Greenhouse-Geisser correction) revealed a significant Group*Structure interaction, \(F(3.06, 79.75) = 2.71, p = .05, \eta_{p}^2 = .17\), no effect of Structure, \(F(3.06, 79.75) = 2.05, p = .11\), and no effect of Group, \(F(1, 24) = .11, p = .74\). Post-hoc comparisons between groups revealed that Experimental participants significantly outperformed Controls on structure A, \(t(24) = 2.44, p = .02, d = .92\), and B, \(t(24) = 2.18, p = .03, d = .30\), but underperformed Controls on structure C, and that difference approached significance, \(t(24) = 2.03, p = .054, d = .80\). There were no between group differences on structures D, E, or F, all \(ps > .34\). The Experimental group outperformed the Control group on structures A and B in the expected direction, but underperformed Controls in the opposite direction on structure C.
To further investigate the pattern of performance across the different syntactic structures in the grammaticality judgment task, Bonferroni-adjusted post-hoc pairwise comparisons were conducted. The performance of the Experimental group across the different structures is reported in Table 3.3 below. Experimental participants differed in their overall endorsement of structure pairs A(TSPOV) and C (TVSPO), \( p = .04 \), A and E (TSVPO), \( p = .01 \), and A and F (TBSOP), \( p = .03 \). No other pair-wise comparisons were significant.
Table 3.3. Bonferroni-adjusted pairwise comparisons between performance on the different structures in the grammaticality judgment task in the Experimental group.

<table>
<thead>
<tr>
<th>(I) Structure</th>
<th>(J) Structure</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.044*</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>.011*</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>.034*</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.08</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.00</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.00</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Overall, the results suggest some evidence of a small amount of learning in the Experimental group; however, learning was only found in structures A and B. Therefore, there is some weak evidence for statistical learning of L2 syntax under incidental learning conditions. However, the extent to which this learning was the result of learning mechanisms that performed statistical computations must be addressed before any strong conclusions can be made with regard to the actual mechanisms of learning found in this experiment.
3.5.2. Research question #2

The second research question asked whether there would be any evidence that adult learners perform statistical computations on their input. To investigate this possibility, analyses were conducted on the Experimental group’s reading times from the exposure phase. Although there was no overall change in the Experimental group’s reading times during the exposure phase, those data were not coded differently for high and low probability sequences. Therefore, it remains an open question how Experimental participants reacted to changes in probability.

Reading times were analyzed in order to determine whether participants were computing statistics during the learning phase. Hypothesis 2 predicted faster reading times for high-probability over low-probability transitions and for sentences with higher mean transitional probabilities over sentences with lower mean transitional probability. In order to evaluate this prediction, two analyses were conducted to see if Experimental participants were computing statistics. The first analysis was conducted on reading times for high- and low-probability transitions. Reading times for each syntactic category bigram sequence were averaged across participants by block. A 2x4 repeated measures ANOVA was run with Probability (High v. Low) and Block (1, 2, 3, 4) as within-subjects variables and reading times as the dependent variable. The ANOVA revealed no significant effects of Probability, $F(1, 12) = .67, p = .44$, or Block, $F(2, 24) = 2.37, p = .12$, and no interaction, $F(2, 24) = 3.11, p = .08$. Thus, the results show no reliable reading time decreases for High Probability transitions over Low Probability transitions, which is inconsistent with the prediction from hypothesis 2 that participants would compute statistics between syntactic categories.

The second analysis assessed whether the mean transitional probabilities of each experimental sentence as a whole would result in different mean reading times for each structure.
(Figure 3.4 below). To assess this question, a 3x4 repeated measures ANOVA was performed with Structure (A, B, and C; transitional probabilities, A: $M_{TP} = .58$, B: $M_{TP} = .42$, C: $M_{TP} = .50$) and Block (1, 2, 3, 4) as within-subjects variables. The dependent measure was mean reading times per sentence structure per block across participants. Analyses revealed a main effect of Block, $F(3, 36) = 4.38$, $p = .01$, $\eta^2_p = .27$. There was also a main effect of Structure, $F(2, 24) = 4.11$, $p < .05$, $\eta^2_p = .26$. There was no Block*Structure interaction, $p = .24$, which probably reflects the general trend toward of a reading time speedup for all structures.

![Figure 3.4](image.jpg)

Figure 3.4. Mean reading times for the different exposure phase structures in the Experimental group by block.

Now, it is unlikely that the significant differences between reading times was due to the mean transitional probabilities that each structure contained. Such a finding would predict, for
example, that the largest speedup in reading time would occur for the highest probability
structure (A [TSPOV]). However, the largest speedup across the mean reading times of each
block was found on structure C, which had the middle mean transitional probability of 0.50.
Likewise, the mean reading times in Block 4 (by which time one might have expected a
statistical learning effect) show that participants were faster at processing structures A and B
than structure C, which is inconsistent with what would be expected if participants were
computing the mean transitional probabilities of those structures. In other words, it did not
appear that participants’ sensitivity to structure was based in the statistics of the input.

Overall, the combined reading time data show no evidence of any influence of statistical
learning on reading times during the exposure phase, contrary to hypothesis 2. Experimental
participants did not get faster at processing high transitional probability sequences than low
transitional probability sequences. Moreover, analyses of reading times indicated that the
Experimental group processed the three sentence structures differently, but not clearly in
accordance with their statistical structure.

3.5.3. Summary of the behavioral results of Experiment 1
The present experiment found no evidence of statistical learning of L2 syntax under incidental
conditions by reading time measures. However, there was some evidence of learning as
measured by generalization to a grammaticality judgment task consisting of new sentences.
Classification performance on the grammaticality judgment task revealed that Experimental
participants were only able to perform above Controls and chance on grammatical structures A
and B. On grammatical structure C they scored below Controls and not differently from chance,
and were not significantly different from controls or chance on the ungrammatical structures D,
E, and F. What led to this pattern of results? One simple, but plausible, explanation is that structures A and B were more easily learned due to their common structure. Structures A and B share the same syntactic categories at the edges of sentences (i.e., TSPOV and TSOPV). The edges of stimuli are known to attract attentional processing (e.g., Endress, Carden, Versace, & Hauser, 2009). If learners merely tracked these edge categories—or even just learned that the verb can go at the end of sentences—the pattern of performance could have been reproduced. That is, verb finality was likely confounded with grammaticality.

The aim of the computational simulations reported in the following section was to address research question #3. If hypothesis 2 were supported, and the adults in Experiment 1 appeared to compute statistics, then it would be plausible to expect the SRN to be able to replicate their behavior (hypothesis 3). However, since hypothesis 2 was not supported, the remaining prediction from hypothesis 4 was that PARSER would be better at capturing the human pattern of performance across the different test structures.

### 3.5.4. Computational Simulations

Experiment 1 showed evidence that adults could learn aspects of L2 syntax on the basis of input containing probabilistic cues under incidental conditions. However, the results showed no evidence of statistical computations across those probabilistic cues, leaving open the question of what was learned in the Experimental group. The aim of the present computational simulations was to investigate whether the pattern of performance for the Experimental group across the structures in the grammaticality judgment task is better replicated by a SRN or PARSER. If the SRN better models the learning found in the human experiment, then it implies that human sensitivity to the probabilistic structure of the syntax may have been derived from implicit
predictive computations analogous to those of the SRN (despite being unable to find evidence for such mechanisms from the behavioral measures). On the other hand, if PARSER better models the learning found in the present human experiment, then it implies that analogous mechanisms may be at work in humans. What are these mechanisms? They are the dynamic interaction of chunk formation and general principles of associative learning and memory, especially memory decay and interference.

3.5.4.1. Stimuli

Both the SRN and PARSER were trained on the same sentence structure templates as the human participants. In other words, the SRN and PARSER were trained on grammatical syntactic phrase sequences. Both models were exposed to each syntactic structure 32 times in a random order, just like human participants. The input coding for each sentence is shown in Table 3.4 below. In the coding scheme, letters represent syntactic phrase categories. An asterisk [*] represents the beginning of a sentence and a pound sign [#] represents the end of a sentence. The lexical items of sentences were not input to either model. Instead, the coding system assumed that syntactic phrases (or their functional equivalents) are processing primitives. This is done for two reasons. First, it provides a measure of comparability with previous studies on mechanisms involved in semiartificial language learning (e.g., Williams, 2010, Experiment 2). Second, as has been pointed out elsewhere (e.g., Chang, Dell, & Bock, 2006; Williams & Kuribara, 2008), it is fairly safe to assume that adult L2 learners have recourse to some knowledge of abstract linguistic categories and apply this knowledge to their L2s. Since the participants in the present experiment were able to generalize to new sentences, such a stimuli

32 This coding scheme was adopted for comparability with Williams’ (2010; Williams & Kuribara, 2008) coding of the SRN. The same simulations were done without explicitly coding the beginnings and ends of the sequences and the results were essentially the same as those reported here.
coding measure is also necessary. This is not inherently a problem for the SRN or PARSER. For example, the perceptual primitives that PARSER uses as its units can be abstract (Perruchet & Vinter, 2002).

Table 3.4. Stimuli coding scheme for the SRN and Parser.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Syntactic</td>
<td>TEMP PHRASE – SUBJ – PREP PHRASE – OBJ – VERB PHRASE</td>
</tr>
<tr>
<td>Structure</td>
<td>* T S P O V #</td>
</tr>
<tr>
<td>Coding</td>
<td></td>
</tr>
<tr>
<td>B Syntactic</td>
<td>TEMP PHRASE – SUBJ – OBJ – PREP PHRASE – VERB PHRASE</td>
</tr>
<tr>
<td>Structure</td>
<td>* T S O P V #</td>
</tr>
<tr>
<td>Coding</td>
<td></td>
</tr>
<tr>
<td>C Syntactic</td>
<td>TEMP PHRASE – VERB PHRASE – SUBJ – PREP PHRASE – OBJ</td>
</tr>
<tr>
<td>Structure</td>
<td>* T V S P O #</td>
</tr>
<tr>
<td>Coding</td>
<td></td>
</tr>
</tbody>
</table>

3.5.4.2. Parameters

Following Boucher and Dienes (2003) both the SRN and PARSER were trained across a variety of parameters. This increases the likelihood that either model’s fit to the human data would be due to intrinsic properties of the model rather than to idiosyncratic, specific parameters that just happen to match human behavior.

Simple recurrent network. Simulated participants (matched for the number of human participants, N=13) were “exposed” to the 96 training sentences presented in random order (i.e., 32 exposures to each syntactic structure). To test the SRN after learning, the activation of the target output node (the correct next syntactic phrase in a sequence) was recorded as a proportion of the activation of all output nodes. This is known as the Luce ratio, and it serves to measure the SRN’s accuracy at predicting the next item in a sequence (for an explanation of the Luce ratio
scoring procedure, see Williams & Kuribara, 2008). These values were then averaged over each test sentence to provide an index of learning in the network.

Table 3.5. Parameter range for the SRN. Only variable parameters are listed, otherwise all parameters were set to default values specified in Ruh and Westermann (2009).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Upper bound</th>
<th>Lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of Hidden and Context Units</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Momentum</td>
<td>0.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

At the beginning of each simulation, the SRN contained seven localist input and output nodes (i.e., one node per syntactic phrase category plus the beginning and ending nodes) and randomly selected connection weights between the nodes. Each simulated SRN participant consisted of a single set of parameters randomly chosen from the range of parameters shown in Table 3.5. To ensure that a given SRN participant’s performance was not due to the initial state of the network (i.e., the initial connection weights), the performance of each simulated SRN participant was calculated as the average of five individual simulations (“runs”) using that SRN participant’s chosen parameters. For each run, the SRN participant was reset to an initial state with connection weights selected randomly between [-.5 .5]. After five complete runs, the results for that set of parameters (i.e., that SRN participant) were averaged into a score for that SRN participant. In other words, performance of each simulated participant was actually the average performance of a single set of SRN parameters (number of hidden and context units) over five independent simulations, each with random initial connection weights. This process was completed for each of the thirteen simulated participants.
PARSER. As with the SRN, the simulated PARSER participants (N = 13) were each given a random set of parameters within the ranges listed in Table 3.6 below. Also as with the SRN, the score for each simulated PARSER participant was the average of that participant over five runs. In other words, each PARSER participant was simulated five times with the same parameters but different initial settings. To measure learning, the weights of whole syntactic structures (e.g., the unit *TSPOV# in PARSER’s memory) were taken as scores. When whole syntactic structures did not exist in PARSER’s memory, chunks that did exist in PARSER’s memory that had appeared in the grammatical syntactic structures had their weights summed and then averaged across runs to form a single participant’s average performance. For example, TSP would count as a chunk from structure A, TSPOV, as would V#. When this occurred, their chunk weights would be summed to yield a chunk strength for structure A. This scoring procedure simply reflects the fact that participants need not have learned entire syntactic structures to endorse sentences, but could have performed exclusively on the basis of fragmentary chunk knowledge.

Table 3.6. Parameter range for PARSER. Only variable parameters are listed; otherwise all parameters were set to default values specified in Perruchet and Vinter (1998).

<table>
<thead>
<tr>
<th></th>
<th>Decay</th>
<th>Interference</th>
<th>N of Perceptual Primitives</th>
<th>Initial Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper bound</td>
<td>0.10</td>
<td>0.020</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>Lower bound</td>
<td>0.02</td>
<td>0.005</td>
<td>1</td>
<td>0.75</td>
</tr>
</tbody>
</table>
3.5.4.3. Computational simulation results

The results of the simulations are reported in Table 3.7 below. Simulation data from the SRN and PARSER were analyzed in order to determine the extent to which either model was able to learn the three target syntactic structures.

Table 3.7. Mean Luce ratios (SRN) and chunk weights (PARSER).

<table>
<thead>
<tr>
<th></th>
<th>Grammatical</th>
<th>Ungrammatical</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>SRN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.528</td>
<td>.502</td>
<td>.512</td>
</tr>
<tr>
<td>SD</td>
<td>.124</td>
<td>.103</td>
<td>.115</td>
</tr>
<tr>
<td>SE</td>
<td>.034</td>
<td>.028</td>
<td>.032</td>
</tr>
<tr>
<td>PARSER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>26.23</td>
<td>23.58</td>
<td>17.18</td>
</tr>
<tr>
<td>SD</td>
<td>12.42</td>
<td>10.98</td>
<td>7.93</td>
</tr>
<tr>
<td>SE</td>
<td>3.44</td>
<td>3.04</td>
<td>2.20</td>
</tr>
</tbody>
</table>

For the SRN, a repeated measures ANOVA (with Greenhouse-Geisser correction) on Luce ratio scores with Structure (A, B, C, D, E, F) as within-subjects factor revealed a significant effect of Structure, $F(1.24, 14.89) = 47.70, p < .001, \eta^2_p = .79$. Moreover, the SRN showed significantly higher Luce values for grammatical items than ungrammatical items, $t(12) = 7.43, p < .001, d = 1.22$. The results indicate that the SRN learned to discriminate grammatical from ungrammatical structures.

In PARSER, 9/13 simulated participants were actually able to learn A, B, and C as whole templates. The other 4 were only partially able to do so, and required the chunk weight scoring procedure outlined above. Chunk weights were mandatorily used to compute scores for ungrammatical structures in PARSER, since PARSER could not have an ungrammatical whole structure in its memory. A repeated measures ANOVA (with Greenhouse-Geisser correction) on chunk weights in PARSER with Structure (A, B, C, D, E, F) as within-subjects factor revealed a
main effect of Structure, $F(2.43, 29.23) = 40.42, p < .001, \eta_p^2 = .77$. Moreover, chunk weight values were significantly larger for grammatical items than ungrammatical items, $t(12) = 25.54, p < .001, d = 0.94$. Taken together, these results demonstrate that both the SRN and PARSER were actually better at discriminating grammatical and ungrammatical items than human adults. Possible reasons for the superior performance by the models are discussed below.

To investigate whether either model could better simulate the human pattern of performance, though, further analyses were conducted on the significant effect of Structure. Bonferroni-adjusted post-hoc pairwise comparisons were conducted on performance differences between grammatical and ungrammatical structures in the SRN and PARSER (Table 3.8 below). The same pairwise comparisons that were conducted on the Experimental group showed significant differences in performance on structure pairs A-C, A-E, and A-F, while the rest of the comparisons were non-significant. Pairwise comparisons on the simulated performance of the SRN and PARSER showed that PARSER was better able to account for human performance patterns across the different syntactic structures than the SRN, accounting for 9/14 of the pairwise comparisons in the Experimental group, while the SRN only accounted for 5/14 of the comparisons. Thus, PARSER was better able to reproduce the Experimental group’s pattern of performance on the different structures in the grammaticality judgment task.
Table 3.8. Bonferroni-adjusted post-hoc pairwise comparisons of patterns of performance across the different structures in the grammaticality judgment task for Humans (Experimental group), the SRN, and PARSER. Comparisons in which the computational models match the results of the Experimental group are denoted by an asterisk.

<table>
<thead>
<tr>
<th>(I) Structure</th>
<th>(J) Structure</th>
<th>HUMAN Sig. a</th>
<th>SRN Sig. a</th>
<th>PARSER Sig. a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>1.000</td>
<td>0.032</td>
<td>1.000*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.044</td>
<td>0.002*</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.298</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>0.011</td>
<td>0.001*</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.034</td>
<td>0.000*</td>
<td>0.000*</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>0.159</td>
<td>0.643*</td>
<td>0.102*</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>0.272</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.082</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000*</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.000</td>
<td>0.001</td>
<td>1.000*</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>1.000</td>
<td>0.052</td>
<td>0.618*</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.000</td>
<td>0.001</td>
<td>0.067*</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>1.000</td>
<td>0.411*</td>
<td>0.005</td>
</tr>
</tbody>
</table>

3.6. Discussion

Experiment 1 was designed to answer three research questions. The first was to establish whether there was any evidence of statistical learning of L2 syntax under incidental learning conditions. There was evidence of a modest learning effect in the Experimental group’s performance on the grammaticality judgment task, but only on structures A and B. The Experimental group performed worse on structure C than the Control group, and was not significantly different in
accuracy on ungrammatical items. Moreover, there was no evidence of learning by any reaction time measures. Taken together, the results suggest some evidence for learning L2 syntax under incidental conditions. The second research question asked whether there would be any behavioral evidence that adult learners perform statistical computations on their input. There was no behavioral evidence of statistical computation. That is, reading time speedup in the Experimental group showed no evidence of statistical computation.

The third research question asked whether the human data could better be replicated by a learning mechanism that computes predictive statistics (SRN) or a learning mechanism that forms increasingly complex chunks (PARSER). It was shown that both models substantially outperformed humans; however, PARSER was better able to replicate the pattern of human performance across the different structures in the grammaticality judgment task. This finding suggests that chunk formation processes may have been involved in the small, but significant, learning effects in the behavioral experiment. Importantly, the finding of an advantage for chunk-based models over connectionist models is consistent with previous research in different domains, such as speech segmentation (e.g., Giroux & Rey, 2009; Perruchet & Peereman, 2004; Perruchet & Vinter, 1998) and artificial grammar learning (e.g., Boucher & Dienes, 2003).

On the other hand, the present findings appear to contrast with previous studies on syntactic development and semiartificial language learning, which found that SRNs can simulate many aspects of human behavioral data (Williams, 2010; Williams & Kuribara, 2008). Why did the present study not reproduce this finding? This remains an open question worthy of further investigation, but there is one possible explanation. It could simply be that the SRN and PARSER simulate two different learning mechanisms, both of which are available to humans and might be more or less likely to operate on different types of stimuli. If this were the case, we
would expect PARSER and SRNs to be able to capture human data to different extents depending on which mechanism was involved. On this view, it may be that the chunk-based learning found in the present study was a result of, say, the relative simplicity and stability of the present semiartificial language stimuli, especially considering that 2/3 of the input sentences had the form TSxxV. However, keep in mind that Williams (2010) found that the SRN was best able to fit the human data when humans were trained on meaningless syntactic category analogues (96% shared variance). When the SRN was compared with performance on the actual semiartificial language learning (as was done here), the shared variance dropped to 40% and 66%. Thus, the inability of the SRN to fit the human data here may not be so different from the results Williams obtained when comparing it to human performance.

In terms of theories and models of SLA, the present findings support accounts wherein the earliest phases of syntactic development are memory-based and rely on simple associative learning mechanisms and general cognitive capacities (e.g., Ettlinger et al., in submission; Perruchet & Tillman, 2010; Ullman, 2004, 2005). The results are also consistent with usage-based accounts of SLA, which posit that L2 syntactic development involves the learning of increasingly well-formed abstract syntactic chunks (i.e., constructions; N. Ellis, 2005, 2006, 2013; N. Ellis & Larsen-Freeman, 2009). However, the use of semiartificial language stimuli means that these interpretations should be treated with caution, since combining phrase-level syntactic categories from one language with larger syntactic constructions from another language may limit the possibility of finding learning effects based on language-specific learning mechanisms.

Another limitation of this study is that the use of the plausibility judgment task meant that half the exposure phase stimuli (48 sentences) were semantically anomalous. It is not clear how
semantically implausible sentences may have adversely influenced learning; however, there is one very important limitation that is immediately apparent. Up to now, it has been repeatedly noted that the syntactic structures contained a frequent pattern: TSxxV for 67% of sentences. Given that the final word of implausible sentences was what made them semantically anomalous, it is possible that these verb phrases and object phrases that ended sentences made them more salient. If so, this would likely influence the Experimental participants to focus on the high-frequency verb-final pattern. Therefore, it is important in future research to avoid exposure phase tasks that might so overtly direct participants’ attention to certain patterns in the stimuli.

Finally, because four participants had to be left out of the analysis, the sample size for this experiment was small. In the future, it will be important to use larger samples sizes that have been established on the basis of a priori power analyses. Only then will it be clear to what extent the present results—especially the non-significant ones—can be regarded as generalizable.

In sum, the results of Experiment 1 suggest that some incidental learning of L2 syntax is possible, but it is limited, at least in the present circumstances. Participants were able to acquire knowledge that enabled them to accurately classify two of the three training structures. Computational simulations suggested that the knowledge they acquired took the form of chunks, although the abstractness of those chunks remains an open question. Abstractness aside, chunks are typically conceived of as units of knowledge in declarative memory (e.g., N. Ellis, 1998, 2005; Hagoort, 2005; Ullman, 2004, 2005) and are often, but not always, considered to be conscious (e.g., Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990). However, the conscious status of the resulting knowledge was not assessed in the present experiment. There was also no assessment of at what point the syntactic knowledge was acquired or how it
developed throughout training. The experiment reported in the next chapter aims to address these aspects of the development of L2 syntax under incidental learning conditions.
Chapter IV: Experiment 2

4.1. Introduction
The results of Experiment 1 provided some evidence for a role for chunk formation processes in early L2 syntax development. In terms of the representation of L2 syntax, this finding is consistent with usage-based accounts of SLA which posit that L2 syntax is construction-based (e.g., N. Ellis, 2005, 2008). Moreover, the fact that PARSER provided a better overall fit to the human data than the SRN suggests that simple chunk formation processes rooted in the natural properties of awareness and memory, which are crucial for the operation of PARSER, may have been involved in participants’ performance. These conclusions are consistent with the possibility that learners acquire conscious knowledge of L2 syntax under incidental learning conditions. Moreover, these conclusions raise a host of new questions. For example, how does conscious knowledge of L2 syntax develop over time in the early phases of incidental learning? The aim of Experiment 2 was to further investigate four related aspects of this larger question. The four research questions are previewed below, but are discussed in more detail along with their hypotheses and rationales in section 4.1.2.

1. At what point during early incidental learning do learners initially develop conscious knowledge of L2 syntax?
2. Do learners acquire conscious knowledge of L2 syntax in the form of declarative memories, microrules, and/or changes in subjective fluency?
3. Is declarative memory for L2 syntax passively memorized throughout training?
4. Does declarative memory for L2 syntax from early in training deteriorate during the course of training?

In order to answer these questions, Experiment 2 used a similar methodology as Experiment 1, but with the addition of three measures of conscious knowledge, which were combined in a novel way in order to triangulate different types of conscious knowledge. In the following sections, the measures of conscious knowledge are introduced first because understanding them is a prerequisite for understanding the hypotheses stated in the subsequent section. I then proceed to describe methods and results of Experiment 2.

4.1.1. Measures of conscious knowledge

In order to investigate the development of conscious knowledge of L2 syntax under incidental learning conditions, measures of conscious knowledge were introduced to the experimental design of Experiment 1. Three measures were used in order to triangulate conscious knowledge at various levels. First, a recognition memory test was used to assess conscious knowledge in the form of declarative memory (e.g., Medina, 2008; Shanks, 2005; Wixted & Squire, 2010).

Second, a retrospective verbal report interview was used to assess learners’ higher-order explicit knowledge (e.g., Rebuschat, 2008; Williams, 2005). Finally, a method new to SLA, on-line subjective fluency ratings were used to assess changes in lower-order conscious knowledge during the exposure phase. These measures are explained in turn below.

4.1.1.1. Recognition memory test

Researchers interested in the relationship between awareness, learning, and memory typically employ recognition memory tests as objective tests of conscious knowledge (e.g., Perruchet et
In a recognition memory test, participants are asked to discriminate test items by indicating whether or not they have seen them before (either a binary yes-no choice or on a Likert-type scale). Participants are considered to have recognition memory if they can correctly discriminate between stimuli they have seen during training (old items) and stimuli they have not seen during training (new items).

What kinds of conscious knowledge do recognition memory tests measure? Performance on a recognition memory test is thought to be mediated by one or more component processes: recollection, familiarity, and/or priming. Recollection is when performance on the recognition memory test is driven by actual recall of the previous exposure to the stimulus (e.g., an episodic memory for a training stimulus). Familiarity is a single strength-of-evidence memory signal which indexes the similarity between a previously seen stimulus and a current stimulus. Although various models posit different roles for recollection and familiarity in recognition memory, nearly all these models agree that familiarity is a vital component which can drive accurate performance in a recognition memory test without actual recollection. That is, most models of recognition memory agree that participants can accurately discriminate old and new stimuli without actually recalling the study context details and without priming (e.g., Berry et al., 2008; Shanks & Berry, 2012; Wixted & Squire, 2010; Yonelinas, 2002). Presently, both recollection and familiarity are considered conscious forms of knowledge (e.g., Wixted & Squire, 2010).

However, implicit knowledge may “contaminate” recognition memory measures (Reingold & Merikle, 1988). For example, Shanks and Johnstone (1999) argue that recognition memory measures may index the influence of implicit knowledge because with increased practice, participants’ implicit mechanisms may be able to process information faster.
(henceforth, priming). This faster processing may lead participants to have the subjective feeling that the sentence is old because it was easy to process. To assess whether there is any influence of implicit priming within an experimental setting, Shanks and Johnstone (1999) recommend measuring reaction time latencies for recognition memory test items that participants judge to be old and new, regardless of the objective status of those test items as old and new. If reaction times are significantly faster for test items that participants judge to be old than items that participants judge to be new, then priming may have influenced recognition memory performance, i.e., because faster reaction times may have led to a feeling of “oldness.” If no such priming effect is present, though, then it suggests that implicit knowledge did not influence recognition memory performance, at least inasmuch as implicit knowledge manifests itself in the form of priming.

Finally, it is possible that recognition memory may be contaminated by semantic knowledge that is not driven by recollection, familiarity, or priming. During a training phase, participants may form higher-order explicit knowledge (i.e., microrules) and use these in the recognition memory test. Assuming their higher-order explicit knowledge is accurate, this would enable participants to accurately discriminate Old and New items in the recognition memory test without having any of the knowledge typically believed to be responsible for recognition memory performance. For example, a native-speaker of English could accurately reject some novel sentences of English purely on the basis of metalinguistic knowledge about what is likely to occur in English, e.g., *The snowman licked the copy of Studies in Second Language Acquisition.*
Operationalization of conscious knowledge in the recognition memory test

Conscious knowledge, as assessed by a recognition memory test, is operationalized as the ability to discriminate Old from New stimuli when the influences of implicit priming and higher-order explicit knowledge have been partialed out. This conscious knowledge is assumed to reflect, at least, a familiarity-based declarative memory. In order to account for the possibility of contamination from implicit priming, the procedure outlined by Shanks and Johnstone (1999) is applied to the recognition memory test and reaction time data. In order to account for the possibility of contamination on the recognition memory test by higher-order explicit knowledge, statistical analyses will examine whether there are any main effects of Old-New discrimination above and beyond any main effects of higher-order explicit knowledge.

4.1.1.2. Retrospective verbal report interview

Retrospective verbal reports are taken as measures of higher-order explicit knowledge. Knowledge is considered conscious in the retrospective verbal report if participants verbalize (1) complete metalinguistic rules, (2) partially-accurate metalinguistic knowledge (microrules), or (3) verbalizable memory for exemplars. In other words, conscious knowledge in the retrospective verbal reports is accurate, albeit possibly incomplete, metalinguistic knowledge. Typically, knowledge is considered to be implicit if participants cannot verbalize the knowledge that supports their accurate performance. However, it is probable that conscious knowledge exists below the verbalizable threshold (e.g., Shanks & St. John, 1994). For this reason, retrospective verbal reports have been criticized for being insensitive measures of conscious knowledge (e.g., Shanks & St. John, 1994). For one, participants may not verbalize their knowledge because they have low confidence in it (Rebuschat, 2008), because it is difficult to verbalize, or because they
believe it to be irrelevant for their performance. Consequently, participants may have conscious knowledge which they simply do not or cannot report. Likewise, even when verbal reports are used on-line (i.e., think-alouds, e.g., Leow, 1997) or in immediate off-line verbal reports (e.g., Philp, 2003), it may be that the low-level contents of awareness are so inherently fleeting and difficult to verbalize that they go unreported. Thus, it may be best to view verbal reports as reflecting higher levels of awareness: conscious knowledge whose representation is strong enough or stable enough to allow for verbalization (Cleeremans, 2007).

Operationalization of conscious knowledge in the retrospective verbal report interviews

Because Experiment 2 is focused on conscious knowledge—rather than awareness at the point of learning—retrospective verbal reports are used. Participants are given a retrospective verbal report interview immediately following the recognition memory test. The interview takes place after the test (rather than during it) to avoid possible on-line reactivity effects (cf. Goo, 2010; Rebuschat et al., in prep; Sachs & Polio, 2007; Sanz et al., 2009). Conscious knowledge, as assessed by retrospective verbal reports, is operationalized as the ability to produce partial or complete metalinguistic information. Although it is an extreme example, participants may be able to verbalize complete metalinguistic rules for certain training sentences (e.g., “Transitional probabilities of 67% and 33% existed between syntactic categories,” or “Sentences consisted of TIME MARKER, NOUN, PREPOSITIONAL PHRASE, OBJECT, and then VERB”). Alternatively, participants may be able to produce microrules (e.g., “Verbs come at the end a lot,” “The doer of the action was in the second position,” or “The subject of the sentence was usually near the beginning.”). Finally, note that for present purposes metalinguistic information need not be abstract. Participants could, in principle, produce sentences or fragments of
sentences (e.g., “I remember reading ‘Yesterday Charlie at the supermarket…’ but I don’t remember the rest of the sentence). This admittedly broad definition of metalinguistic information is preferred, since participants often have conscious knowledge but no knowledge of metalinguistic jargon.

4.1.1.3. Subjective fluency ratings

Subjective fluency ratings were designed to assess the trial-by-trial development of lower-order conscious knowledge in the form of changes in subjective experience of fluency. The logic behind the subjective fluency ratings is derived from an influential model of how learning in incidental conditions leads to a change in the subjective experience of the learner. Servan-Schreiber and Anderson (1990) propose that exposure leads to improved perceptual processing (increased fluency), which alters subjective experience of the stimuli. Thus, more fluent processing (i.e., faster reading of sentences during the exposure phase) may lead participants to have a change in conscious experience: a conscious experience that the stimuli are easier to parse, etc. Thus, the subjective fluency ratings, which, to the author’s knowledge, are new online measures of lower-order conscious knowledge in SLA, are intended to tap such a change in learners’ low-level subjective experience of the stimuli. If learners become more fluent at processing the training stimuli in their processing speed, comparable changes to their subjective experience of reading may occur.

Operationalization of conscious knowledge using subjective fluency ratings

Following Servan-Schreiber and Anderson’s (1990) hypothesis, lower-order explicit knowledge, as measured by subjective fluency ratings, can be straightforwardly operationalized as an
increase in participants’ subjective ease of reading sentences in the exposure phase. On the other hand, it is not as easy to interpret a finding which shows increases in participants’ subjective difficulty of reading sentences in the exposure phase. One possibility is that increases in subjective difficulty of reading may indicate ongoing organizational processes that have not yet found an optimal solution (similar perhaps to the U-shaped learning curve in language acquisition). Alternatively, increases in subjective difficulty of reading may indicate processing difficulties associated with unexpected events (e.g., Haider & Frensch, 2009).

4.1.2. Research questions, hypotheses, and operationalizations

Research question #1: At what point during early incidental learning do learners initially develop conscious knowledge of L2 syntax?

Hypothesis 1: Experimental participants will acquire conscious knowledge of L2 syntax as early as the briefest training period (18 exposures). Consistent with previous research on the development of conscious knowledge under incidental learning conditons, conscious knowledge is predicted to consist of successful discrimination of Old and New items in the recognition memory test (e.g., Perruchet et al., 1997) and verbalizable microrules (e.g., Haider & Frensch, 2005; Rose et al., 2010). However, due to the brevity of the 18-exposure condition, participants are not expected to have developed conscious knowledge by subjective fluency rating measures, consistent with previous suggestions that fluency changes are neither rapid nor large in their effects (e.g., Perruchet, 2008).

The current hypothesis follows from the conclusions of studies of incidental learning that relied primarily on the use of non-linguistic stimuli. It is possible that the rapid development of conscious knowledge of linguistic stimuli is different. The results of incidental learning studies
in SLA have provided mixed evidence for rapid acquisition of conscious knowledge under incidental learning conditions. For example, Hamrick and Rebuschat (2012, 2013) demonstrated that adults could learn conscious lexical knowledge under incidental conditions in as little as five minutes. Other studies have suggested that incidental learning of L2 vocabulary has been shown to be a slow process (e.g., Pellicer-Sanchez & Schmitt, 2010). However, many of these studies were based on reading textbooks or novels, and necessarily included longer learning phases, whereas Hamrick and Rebuschat (2012, 2013) investigated rapid word learning in a laboratory setting with no text-reading component. Although participants in Experiment 2 are reading sentences, the incidental learning conditions are much more like those of Hamrick and Rebuschat (2012, 2013) and the non-linguistic studies mentioned above. With its similar experimental design, it is expected that the present experiment will produced evidence for rapid learning of conscious knowledge of L2 syntax.

Research question #2: Do learners acquire conscious knowledge of L2 syntax in the form of declarative memories, microrules, and/or changes in subjective fluency?

Hypothesis 2: Experimental participants will acquire conscious knowledge in the form of higher-order explicit knowledge (e.g., microrules), regardless of the duration of their exposure phases. It is empirically well-established that learning under incidental conditions produces microrules (e.g., Dulany et al., 1984, 1985; Hama & Leow, 2010; Rebuschat, 2008; Rebuschat et al., forthcoming). However, to the author’s knowledge, there are not prevalent theoretical accounts of why this so (although see Haider & Frensch, 2005, for an extended discussion of possible explanations). One possibility is that higher-order explicit knowledge is recruited when implicit mechanisms fail to successfully process language, as they often do in L2 learning (N.
Ellis, 2005, p. 308). It is therefore predicted that exposure to the semiartificial language, which will seem like English “gone wrong,” will lead to the development of higher-order explicit knowledge.

**Hypothesis 3:** Experimental participants will acquire conscious knowledge in the form of declarative memories of L2 syntax. Regardless of the duration of their exposure phases (18-, 48-, or 96-exposure), it is predicted that experimental participants will be able to discriminate Old and New items in the recognition memory test. This view is consistent with memory-based approaches to second language acquisition, which predict that learners store complex forms in declarative memory during early phases of L2 syntax acquisition (e.g., Ullman, 2004, 2005). This view is also consistent with usage-based approaches to SLA (e.g., N. Ellis, 2005, 2006, 2008) and usage-based and lexicalist accounts of syntax (e.g., Culicover & Jackendoff, 2006; Goldberg, 2006; Hagoort, 2005; Novick, Kim, & Trueswell, 2003; Vosse & Kempen, 2000), all of which assume that syntax consists of constructions or abstract syntactic chunks (e.g., V-PP) stored in declarative memory.

**Hypothesis 4:** Experimental participants in the 48- and 96-exposure conditions will acquire lower-order explicit knowledge as measured by subjective fluency ratings. Participants in the 18-exposure condition will not acquire such knowledge. This result will be demonstrated by 48- and 96-exposure participants rating sentences during the exposure phase as increasingly easy to read. This hypothesis is grounded in the assumption that learning results in a change to participants’ subjective experience of the stimuli (e.g., Servan-Schreiber & Anderson, 1990).

*Research question #3: Is declarative memory for L2 syntax passively memorized throughout training?*
**Hypothesis 5:** Experimental participants will continuously, and passively, memorized conscious knowledge in the form of declarative memory during the exposure phase. It is therefore expected that Experimental participants will be able to discriminate old vs. new sentences which have been recently presented (i.e., in the final 18 trials of the exposure phase), regardless of the duration of their exposure phase. This expected result is consistent with accounts of SLA and the psychology of learning that argue for a role for the continuous aggregation of conscious knowledge, such as the Fundamental Similarity Hypothesis (Robinson, 1996, 1997), the Self-Organizing Consciousness/PARSER models (Perruchet & Vinter, 1998, 2002), and the Competitive Chunking model (Servan-Schreiber & Anderson, 1990).

**Research question #4:** Does declarative memory for L2 syntax from early in the training phase deteriorate over the course of training?

**Hypothesis 6a:** Experimental participants’ ability to discriminate old vs. new sentences from early in the training phase will deteriorate over the course of training if the declarative memories supporting their behavior are completely bound to the exemplar sentences from the exposure phase. In other words, participants would be expected to forget sentences they have not seen for a while if their knowledge of syntax was limited only to knowledge based on the surface structure of the training sentences. This account assumes that syntax acquisition is exemplar-driven, at least in the early phases. This hypothesis would be supported if participants were significantly better in the recognition memory test at discriminating old vs. new items they had recently seen during the exposure phase than old vs. new items from early in the exposure phase. This possibility is consistent with usage-based accounts of L1 and L2 syntax acquisition (e.g., N. Ellis, 2005, 2008; Tomasello, 2003).
Hypothesis 6b: However, there is another possible outcome. If declarative memory for L2 syntax goes beyond surface structure, then there should be less evidence of decay of that conscious knowledge. Why? Because the abstract syntactic structures remain the same throughout training, thus making all sentences from the training phase the same at the level of abstract syntax. This hypothesis would be supported if participants were not significantly better at discriminating old vs. new items they had recently seen during the exposure phase than old vs. new items from early in the exposure phase. In theory, this result could derive from the fact that at the level of abstract syntax, there are no differences between recently and non-recently presented sentences. This result is consistent with approaches that posit a separation between the cognitive and neural underpinnings of lexical knowledge and syntax, but with syntax still being memory-based (e.g., Ullman, 2004, 2005).

4.2. Methods

4.2.1. Participants

A total of 82 volunteer undergraduates participated in the present experiment (male = 35, female = 47, $M_{age}$ = 19.10, range = 18-24). Participants were randomly assigned to an Experimental or Control group across three conditions, for a total of six independent groups: 18-exposure condition (Exp18 n = 15, Con18 n = 13), 48-exposure condition (Exp48 n = 14, Con48 n = 13), and 96-exposure condition (Exp96 n = 14, Con96 n = 13). Twelve participants reported having another native language in addition to English; these included Spanish (5), Korean (4), Japanese (1), French (1), and Mandarin (1). Sixty-eight reported studying one or more foreign languages, including Spanish (26), French (20), Italian (6), Latin (6), German (5), Mandarin (4), Greek (3), Russian (2), Arabic (1), Catalan (1), and Korean (1). No participants had any prior knowledge of
Persian. Eleven participants were Linguistics majors (1 in the Con18 group, 2 in the Exp48 group, 5 in the Con48 group, and 3 in the Exp96 group). Participants who were enrolled in an Introduction to Linguistics course were offered 5% extra credit on a homework assignment in that course for their participation. None of the Experimental or Control groups differed significantly across age, sex, handedness, linguistics experience, or number of languages spoken (all ps > .05). All participants reported having normal or corrected-to-normal vision and hearing.

4.2.2. Stimuli

Sentences came from the same semiartificial language structures as in Experiment 1. Therefore, the present stimuli maintained the same structural and statistical properties as Experiment 1. However, the sentences were modified in several ways. First, all sentences were made “plausible” so that their compositional meanings depicted possible real world events, since the plausibility judgment task was no longer being used. This manipulation eliminated some undesirable sentence processing effects associated with semantic incongruency, e.g., the drawing of attention to semantically anomalous forms.

Second, a set of core sentences were produced. These sentences consisted of one of seven TEMPORAL PHRASES (over the weekend, last month, this morning, recently, yesterday, this past week, and some time ago), 16 SUBJECT NOUN PHRASES (e.g., the musician, the model, the farmer, the doctor), 32 VERBS (e.g., gave, threw, picked, prescribed), 32 OBJECT NOUN PHRASES (e.g., a concert, a party, some vegetables, some medicine), and 32 PREPOSITIONAL PHRASES (e.g., for charity, for her birthday, on the chalkboard, in the hospital). These core sentences were designed for more consistency between their formal structure (i.e., their syntactic categories) and their semantic structure (i.e., themes, roles, etc.). This was done to avoid
confounds whereby learners do or do not show evidence of learning due to semantic structure as opposed to syntactic structure. However, there was some unavoidable variation, primarily in the prepositional phrases which encoded both Location and Purpose thematic roles. So this limitation, inherent to using semiartificial language stimuli, must be taken into consideration. Crucially, because the core sentences occurred with equal frequency in all three target syntactic structures, there was no consistent relationship between thematic roles and a single syntactic structure.33

Finally, the stimuli had to be modified because of how recognition memory for syntax was measured. In this experiment, recognition memory for syntax was assessed by showing Experimental participants sentences, all of which they had seen before but only half of which occurred in three target syntactic structures from the exposure phase. In order to establish that recognition memory for a sentence was driven by syntax and not just idiosyncratic memory for that sentence, each core sentence appeared in all three target syntactic structures (TSPOV, TSOPV, and TVSPO). In other words, the same core sentences were rotated around each syntactic structure used in the exposure phase. Thus, a given core sentence (e.g., Earlier today the farmer at the market tomatoes sold) was presented once in each of the three syntactic structures (see Figure 4.1 below). No matter which condition participants were in, they would see each core sentence in all three syntactic structures (as before, in random order). For example, during the exposure phase in the 18-exposure condition, participants saw six core sentences three times, once in each of the three syntactic structures, yielding a total of 18 exposures. This manipulation ensured that participants saw the same core sentences the same number of times and the same number of times in each target syntactic structure.

33 While this adds the benefit of superior control over the stimuli, it removes any semantic or pragmatic influence from the syntactic structures themselves. Given increased theoretical emphasis on the meaningful nature of syntactic structures (e.g., Culicover & Jackendoff, 2006; Goldberg, 1995, 2006), this aspect of the stimuli may be a limitation.
As in Experiment 1, Control participants saw the same core sentences as Experimental participants in their respective conditions; however, the syntactic structures of those core sentences were pseudorandomized in two ways. First, the syntactic structures never contained any whole structure repetitions (a given structure, e.g., SPVTO, never occurred twice). Second, syntactic structures that occurred in the recognition memory test were systematically excluded from the pseudorandom set of syntactic structures that the Control participants read during the exposure phase (TSPOV, TSOPV, TVSPO, or any structure from Table 4.1 below).
4.2.2.1. Recognition memory test stimuli

The aim of the recognition memory test was to isolate an effect of syntax on recognition memory, over and above any contributions of lexical and semantic information. To achieve this end, participants were instructed to indicate whether or not they had seen a given sentence before. In other words, they were asked to discriminate Old (previously seen) from New (previously unseen) sentences. Crucially, the terms Old and New are relative in two ways. First, they are relative in the sense that the recognition memory test consisted only of core sentences (i.e., the lexical and compositional semantic content of training sentences) from the exposure phase. The difference between Old and New items is relative to the Experimental groups. Old items followed the three “grammatical” syntactic structures that Experimental participants saw during the exposure phase (TSPOV, TSOPV, and TVSPO), while the New sentences consisted of the lexical and compositional semantic content of previously seen core sentences in novel, but “ungrammatical” configurations (e.g., TPOSV: Earlier today at the market tomatoes the farmer sold). Thus, in the recognition memory task Experimental participants saw old core sentences in syntactic structures that they had seen during the exposure phase (so-called Old items) and old core sentences in syntactic structures that they had not seen during the exposure phases (so-called New items). Lexically and semantically speaking, all items were equally “old.” Therefore, the only information available to participants for discriminating Old from New sentences was whether those sentences had been seen in the exact target syntactic structures from the exposure phase.

The terms Old and New are also relative to the extent that they are specific to the Experimental group. In the recognition memory test, Experimental participants were, in essence, discriminating on the basis of the Old-New status of syntax. Now, the Control participants saw
the same recognition memory test sentences. However, the syntactic structures used in the recognition memory test never occurred in the exposure phase for Control participants. As with the Experimental group, this meant that all test items were equally “old” lexically and semantically. However, syntactically speaking, all test items were “new,” because the Control participants had not seen those syntactic structures before during the exposure phase. Training and testing the Control group in an identical way as the Experimental group, but with the syntactic regularities in the exposure phase removed, improves the likelihood that any evidence of learning in the Experimental group is due to actual learning rather than, say, some source of bias stemming from the experimental materials or procedure (Hamrick & Sachs, in prep).

Importantly, for the sake of simplicity the terms Old and New will be used throughout to refer to the two types of recognition memory test items; however, it is important to keep in mind that these terms are relative.

Recognition memory tests were organized by block. A recognition memory test block consisted of the presentation of three Old sentences and three New sentences. Results from a pilot study suggested that presenting the same core sentence in different syntactic structures in the recognition memory test led to substantial interference effects. For example, presenting participants with *Earlier today the farmer at the market tomatoes sold* (Old, TSPOV) and *Earlier today at the market tomatoes the farmer sold* (New, TPOSV) led to reliable recognition memory accuracy for the first few test items, but that accuracy quickly declined due to increased interference. Therefore, in each recognition memory test block each syntactic structure was assigned only one core sentence, whose lexical and semantic contents were not repeated in any other test sentence. The recognition memory tests in all conditions contained a small number of items. For participants in the 18-exposure condition, the recognition memory test consisted of the

---

34 This is an important limitation in the present design, which is discussed further in limitations section of Chapter 5.
random presentation of six items: three Old sentences (structures TSPOV, TSOPV, and TVSPO each containing different lexical and semantic content) and three New sentences (each containing different lexical and semantic content). New sentences were constructed according to two different sets of syntactic structures in Table 4.1 below (set DEF and set XYZ).

A given participant in the 18-exposure condition would see either New sentences DEF or New sentences XYZ (these were counterbalanced). Participants in the 48- and 96-exposure conditions were asked to discriminate Old-New status for sentences from early in the exposure phase and late in the exposure phase. Early-Old and Early-New items consisted of core sentences from the first 18 trials of the exposure phase. Late-Old and Late-New items consisted of core sentences from the final 18 trials of the exposure phase. Early-New sentences followed the syntactic structures in set DEF from Table 4.1, while Late-New sentences followed the syntactic structures in set XYZ from Table 4.1. In the recognition memory task, participants in the 48- and 96-exposure groups saw Late items in the first test block and Early items in the second test block. This block order was used to ensure that participants in all conditions were always tested first on a sample of sentences from the most recent 18 trials.

In sum, the recognition memory test consisted of one block for 18-exposure participants and two blocks for 48- and 96-exposure participants. Test phase blocks were organized so that comparisons could be made across the different experimental conditions. This allowed for the measurement of recognition memory for syntax at time intervals relevant to the research questions.

- Research question #1: recognition memory after 18 exposures;
- Research question #2: recognition memory after all exposure durations and for Early and Late items;
• *Research question #3:* recognition memory after all exposure durations for Late items and for the Exp18 group;

• *Research question #4:* recognition memory for Early items in the Exp48 and Exp96 groups.

Table 4.1. Syntactic templates used for New (ungrammatical) structures in the recognition memory test.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Syntactic Template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>T P O S V</td>
<td>Earlier today at the market tomatoes the farmer sold.</td>
</tr>
<tr>
<td>E</td>
<td>P T S O V</td>
<td>At the market earlier today the farmer tomatoes sold.</td>
</tr>
<tr>
<td>F</td>
<td>T V O S P</td>
<td>Earlier today sold tomatoes the farmer at the market.</td>
</tr>
<tr>
<td>X</td>
<td>S T V P O</td>
<td>The farmer earlier today sold at the market tomatoes.</td>
</tr>
<tr>
<td>Y</td>
<td>T P S O V</td>
<td>Earlier today at the market the farmer tomatoes sold.</td>
</tr>
<tr>
<td>Z</td>
<td>T V P S O</td>
<td>Earlier today sold at the market the farmer tomatoes.</td>
</tr>
</tbody>
</table>

4.2.3. Procedure

The overall procedure for Experiment 2 was similar to that of Experiment 1, but with a few modifications which are explained in detail below. A schematic of the procedure for Experiment 2 is shown in Figure 4.2 below. Participants were tested individually in a quiet laboratory. They were told that they were participating in a study about reading comprehension under “unusual circumstances.” They were informed that, in their case, “unusual circumstances” meant that they would be reading scrambled sentences.
<table>
<thead>
<tr>
<th>18-Exposure Condition</th>
<th>48-Exposure Condition</th>
<th>96-Exposure Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp18</td>
<td>Con18</td>
<td>Exp48</td>
</tr>
</tbody>
</table>

---

Exposure Phase

- **18 exposures to each of 3 structures**
- **16 exposures to each of 3 structures**
- **48 exposures to each of 3 structures**
- **32 exposures to each of 3 structures**
- **96 exposures to each of 3 structures**

---

Recognition Memory Test

- **6 items:** 3 Old, 3 New
- **12 items:** 3 Old-Late, 3 New-Late, 3 Old-Early, 3 New-Early
- **12 items:** 3 Old-Late, 3 New-Late, 3 Old-Early, 3 New-Early

---

Retrospective Verbal Report Interview

---

Biodata

---

Figure 4.2. Schematic of the procedure for Experimental and Control groups across conditions.

This figure uses the terms Old and New relative to the Experimental group. Section 4.2.2.1 above clarifies the use of the terms Old and New.

**4.2.3.1. Exposure phases**

The experiment was set up as a study on reading fluency. Participants were told that the purpose of the study was to investigate how people read when the reading materials have been scrambled. Participants were told that their task was to read each sentence for meaning as though they were reading a book, article, or blog. They were told that after reading each sentence they would be
asked to indicate how easy or difficult it was to read that sentence on a scale from 1 (very easy) to 6 (very difficult). Figure 4.3 below shows the scale participants used. Each number corresponded to a number placed over re-labeled keyboard keys (i.e., A = 1, D = 2, G = 3, J = 4, L = 5, ‘ = 6). Sentences were presented using the same self-paced non-cumulative moving window design from Experiment 1. Sentences were segmented at syntactic category boundaries, and participants pressed the SPACE BAR to advance through each sentence fragment. Therefore, each SPACE BAR key press corresponded with the presentation of a new syntactic category constituent. The exposure phase consisted of 18, 48, or 96 trials depending on the condition. Each trial consisted of the presentation of a fixation cross, followed by a sentence, and then an ease/difficulty of reading judgment. This judgment served as the subjective fluency rating, an index of participants’ perceptions of their own processing fluency. Sentences were presented in a pseudorandom order (with the first and last 18 trials of the 48-exposure and 96-exposure condition containing the core sentences use in the recognition memory task). Participants were not informed that there would be any kind of test after the exposure phase. The exposure phase took, on average, under five minutes to complete for the 18-exposure condition, ten minutes for the 48-exposure condition, and 20 minutes for the 96-exposure condition.
4.2.3.2. Recognition memory test phase

After the exposure phase, participants were then told that their next task would be to read more sentences, but this time instead of indicating how easy/difficult it was to read each sentence, they would be asked to indicate whether or not they had seen each sentence using another scale from 1 (I have seen this sentence before and I am very sure) to 6 (I have not seen this sentence before and I am very sure), as shown in Figure 4.4. Participants were also told that half of the test sentences would be Old (i.e., exactly the same as the sentences they just read). The other half would be New, i.e., not exactly the same as the sentences they just read (see Appendix for test instructions and stimuli). Sentences were presented using the same self-paced non-cumulative
moving window design from the exposure phase, with each SPACE BAR key press corresponding to the presentation of a new syntactic category constituent. On average, the recognition memory test took five minutes to complete.

Figure 4.4. Recognition memory judgment scale used for test phase trials in all conditions and groups in Experiment 2.

4.2.3.3. Retrospective verbal report interview

After the recognition memory test, participants were engaged in a structured oral interview. They were asked a series of questions, each of which became increasingly targeted at more explicit information. All participants, including Controls, completed the verbal report task. The questions from the verbal report task are reproduced in Table 4.2 below.
Table 4.2. Questions used in the structural oral interview at the end of each participant’s session.

<table>
<thead>
<tr>
<th>Q #</th>
<th>Question</th>
<th>Next Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What were you thinking during the first part of the experiment when you</td>
<td>Go to 2</td>
</tr>
<tr>
<td></td>
<td>were reading the sentences and saying how easy/difficult they were? Did</td>
<td></td>
</tr>
<tr>
<td></td>
<td>you use any strategies or tricks to complete that task?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>In the first part of the experiment, did you feel like the sentences</td>
<td>Go to 3</td>
</tr>
<tr>
<td></td>
<td>overall got easier to read or more difficult or the stayed the same?</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>In the second part of the experiment, we asked you whether or not you</td>
<td>Go to 4</td>
</tr>
<tr>
<td></td>
<td>remembered seeing certain sentences. Did you have any special strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in doing that part?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>What were some of your specific thoughts during that second part?</td>
<td>Go to 5</td>
</tr>
<tr>
<td>5</td>
<td>At any part during the experiment, did you notice any patterns? Did</td>
<td>Go to 6</td>
</tr>
<tr>
<td></td>
<td>anything stand out (especially about the sentences) to you?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Did the sentences seem random or predictable?</td>
<td>If predictable, go to 7; if random, go to 8</td>
</tr>
<tr>
<td>7</td>
<td>What seemed predictable about it?</td>
<td>Ask them to produce an example of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>predictability and ask them when they noticed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the pattern. Then proceed to 9</td>
</tr>
<tr>
<td>8</td>
<td>If I told you that there were actually patterns in those sentences, could</td>
<td>Go to 9</td>
</tr>
<tr>
<td></td>
<td>you say what you think they might be?</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>We actually made the sentences using a set of probabilistic rules. Do you</td>
<td>If they are hesitant, encourage them to share</td>
</tr>
<tr>
<td></td>
<td>think you could describe the rules? Or any rule that you used?</td>
<td>any intuition, vague notion, or suggestion: go</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to 10</td>
</tr>
<tr>
<td>10</td>
<td>If they still don’t mention the right rules and regularities, explain the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>system (i.e., there are 3 word order patterns TSPOV, TSOPV, TVSPO), and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>then ask if they ever considered any of these or parts of these.</td>
<td></td>
</tr>
</tbody>
</table>

4.3. Results

The results of Experiment 2 are broken down as follows. I restate each research hypothesis and report the results from the relevant conditions (i.e., 18-, 48-, and 96-exposure conditions). For each condition, I report only the results of measures of learning pertaining to that research question. Alpha levels were set to 0.05 for all statistical tests. Non-parametric statistical tests were used for comparisons that included any non-normally distributed data. Otherwise, parametric tests were used.
4.3.1. Research question #1

In response to research question #1, it was predicted that Experimental participants would acquire conscious knowledge of L2 syntax by recognition memory and retrospective verbal report measures, but not subjective fluency rating measures, after just 18 exposures (six exposures to each structure). Therefore, only the results from the 18-exposure condition were analyzed. Results from the verbal reports, recognition memory test, and subjective fluency ratings are described in turn.

4.3.1.1. Retrospective verbal report

In order to assess whether participants acquired higher-order explicit knowledge in the 18-exposure phase, the retrospective verbal report interviews from the Exp18 and Con18 groups were analyzed. Analyses of the Exp18 group showed that 14 out of 15 participants were able to verbalize at least some accurate knowledge about the syntactic structure of the exposure phase sentences, but no participants verbalized any knowledge of a whole structure. Therefore, these 14 participants were considered to have acquired higher-order explicit knowledge by the definition laid out in Chapter 1. Only one Control participant reported any knowledge of syntactic structure that would be beneficial at test. This participant (C18_15) reported using a verb-final strategy in the test phase: “I had remembered, or thought I had remembered, a sentence that ended with a verb and so, one of the…several sentences, and so I kind of thought that one.” This participant gave a mean recognition score of 2.33 for Old items and 3.00 for New items, which is at least consistent with what would be expected by using a verb-final microrule. It is unclear why this particular strategy was used, but it is most likely that this participant successfully overgeneralized a memory for a random training string that happened to end in a verb.
Since no other Control participants reported awareness of the target structures, the remainder of the verbal report analyses focuses on the Exp18 group. The 14 aware participants in the Exp18 group expressed different microrules for the syntactic structures in the experiment. However, there were some reliable patterns. For instance, 13 participants reported attending to the placement of the verb, with 11 of these participants specifically indicating that the verb occurred (at least some of the time) at the ends of sentences. A typical report is provided in (1) and (2) below.

(1) *The noun and the verb. The verb kept coming at the- the end…The verb usually came on the last… or the third, I think. I think…sort of…It was predictable, because the verb came last and I know what the nouns do.*

(2) *I started, kind of, to expect, like, verbs to come at the end. You know, I kind of just noticed the patterns, like the waitress bought. No. The waitress…washed.*

Six participants also reported awareness that the temporal phrase came at the beginning of sentences (as in (3) below). Five of the six participants reporting awareness of the temporal phrase position also reported awareness of the verb final pattern.

(3) *First, I wanted to have the, uh, time, when it was done, so that’s what I always looked for at the beginning, before I figured out what else happened. I just, it’s like it was, said like “Yesterday surgery performed” something like that, and then I made sure “Yesterday” was the first thing in my mind and then everything else kind of fell into place.*
Only two participants mentioned the locations of any other syntactic categories, both times being the object phrase OBJECT PHRASE, which participants noted, correctly, could end a sentence or come in the next-to-last position. A typical example is found in (4).

(4) *I noticed the end more ‘cause that’s the last thing I hit...* So there was...um...oh God, help me, my grammar’s awful. *The, uh, object of the verb? Like, I noticed “the book.”*

Thus, the retrospective verbal report interviews indicate that participants in the Exp18 group developed higher-order explicit knowledge after just 18 exposures to the three target syntactic structures, consistent with hypotheses 1.

**4.3.1.2. Recognition memory test**

Hypothesis 1 also predicted that Exp18 participants would develop declarative memories after just 18 exposures, as measured by performance on the recognition memory test. In the recognition memory test, participants were asked to read each test sentence and indicate on a scale from 1 (*I have seen this sentence before and I am very sure*) to 6 (*I have not seen this sentence before and I am very sure*) to what extent they believed that they had seen that sentence before. To investigate whether participants rapidly acquired conscious knowledge of L2 syntax in the form of declarative memories, a recognition memory difference score analysis was conducted.
Table 4.3. Distribution of recognition memory judgments at or below each classification category.

<table>
<thead>
<tr>
<th>Test Item Type</th>
<th>% of Mean Recognition Memory Scores at or Below</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 have seen</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>6.67</td>
</tr>
<tr>
<td>New</td>
<td>0.00</td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>0.00</td>
</tr>
<tr>
<td>New</td>
<td>26.67</td>
</tr>
</tbody>
</table>

Participants in both groups did not tend to use the full recognition memory scale, despite being encouraged to do so in the task instructions. Table 4.3 above indicates that, on average, participants placed 78% of their ratings on the lower end of the recognition memory scale (i.e., the Old side of the scale, 1-3). Since participants did not clearly utilize the lower end of the scale for Old items and the higher end of the scale for New items, it was possible that differences between groups may be obscured. To account for this, difference scores were computed in order to assess how well Experimental and Control participants were able to discriminate Old and New items. Difference scores for each participant were computed as a participant’s mean rating for New test items minus his/her rating for Old test items. These are reported in Table 4.4 below. Bear in mind that the recognition memory rating scale put recognition of the “oldness” of items at lower numerical values and the “newness” of items at higher numerical values on the scale. If one subtracts ratings on Old test items from New test items, then positive difference scores indicate better recognition memory for Old items than New (i.e., implying the presence of recognition memory), and negative difference scores indicate endorsements of New items as

35 Recall that Old and New are relative to the Experimental group. In their exposure phase, Control participants did not see any of the syntactic structures from the recognition memory task.

136
more recognized than Old items (i.e., implying the use of an incorrect memory) and difference scores close to zero imply a lack of an recognition memory. Analysis of difference scores shows that the mean difference scores in the Exp18 group were significantly larger than those in the Con18 group, $t(27) = 3.31, p = .003, d = 1.22$. Thus, analysis of the difference scores suggests that the Experimental group acquired conscious knowledge of L2 syntax after just 18 exposures.

Table 4.4. Descriptive statistics for difference scores by group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Experimental</td>
<td>1.09</td>
<td>3.67</td>
<td>1.01</td>
<td>0.26</td>
<td>0.52</td>
</tr>
<tr>
<td>Control</td>
<td>-0.58</td>
<td>6.67</td>
<td>1.64</td>
<td>0.44</td>
<td>-1.54</td>
</tr>
</tbody>
</table>

Now, one of the aims of the experimental design was to make sure participants rated New items because of their “newness” rather than because of idiosyncratic properties of their syntax, roughly half of the Exp18 (n = 7) and Con18 groups (n = 6) were given New items D, E, and F, and the other half (Exp18 n = 8; Con18 n = 7) were given New items X, Y, and Z. Although this procedure provided a variable pool of New items, it unfortunately meant that comparing differences between groups on the effects of the different structures was only possible with the original group sizes cut in half. For example, comparing groups on structures D, E, and F could only be done for the 7 Exp18 and 6 Con18 participants. The descriptive statistics for the recognition memory test for the Exp18 and Con18 groups are reported. Table 4.5 below show the descriptive statistics for these data, which were normally distributed, $ps > .08$. 
Table 4.5. Descriptive statistics for the recognition memory test in the 18 exposure condition by group and test item type.

<table>
<thead>
<tr>
<th>Group</th>
<th>Test Item Type</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Exp18 - DEF</td>
<td>A</td>
<td>1.62</td>
<td>1.00</td>
<td>0.51</td>
<td>0.18</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.12</td>
<td>2.00</td>
<td>0.64</td>
<td>0.22</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.13</td>
<td>4.00</td>
<td>1.35</td>
<td>0.47</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>2.50</td>
<td>4.00</td>
<td>1.41</td>
<td>0.50</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.00</td>
<td>5.00</td>
<td>1.69</td>
<td>0.59</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>3.75</td>
<td>4.00</td>
<td>1.38</td>
<td>0.49</td>
<td>2.58</td>
</tr>
<tr>
<td>Exp18 - XYZ</td>
<td>A</td>
<td>2.12</td>
<td>4.00</td>
<td>1.80</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.00</td>
<td>4.00</td>
<td>1.60</td>
<td>0.56</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3.50</td>
<td>5.00</td>
<td>1.77</td>
<td>0.62</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>3.50</td>
<td>5.00</td>
<td>2.07</td>
<td>0.73</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>3.75</td>
<td>5.00</td>
<td>2.05</td>
<td>0.72</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>3.62</td>
<td>4.00</td>
<td>1.76</td>
<td>0.62</td>
<td>2.14</td>
</tr>
<tr>
<td>Con18 - DEF</td>
<td>A</td>
<td>2.00</td>
<td>3.00</td>
<td>1.26</td>
<td>0.51</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3.28</td>
<td>5.00</td>
<td>1.79</td>
<td>0.68</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.57</td>
<td>4.00</td>
<td>1.61</td>
<td>0.61</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>4.16</td>
<td>5.00</td>
<td>2.22</td>
<td>0.91</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.14</td>
<td>3.00</td>
<td>1.34</td>
<td>0.51</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.18</td>
<td>4.00</td>
<td>1.57</td>
<td>0.59</td>
<td>0.40</td>
</tr>
<tr>
<td>Con18 - XYZ</td>
<td>A</td>
<td>2.71</td>
<td>5.00</td>
<td>1.60</td>
<td>0.60</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.42</td>
<td>5.00</td>
<td>1.90</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.00</td>
<td>5.00</td>
<td>1.82</td>
<td>0.69</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>1.85</td>
<td>2.00</td>
<td>0.89</td>
<td>0.34</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>1.14</td>
<td>1.00</td>
<td>0.37</td>
<td>0.14</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>1.14</td>
<td>2.00</td>
<td>0.69</td>
<td>0.26</td>
<td>0.50</td>
</tr>
</tbody>
</table>

A qualitative look at the descriptive data shows patterns that are consistent with the difference scores analysis. Regardless of whether the Exp18 participants saw New items D, E, and F or X, Y, and Z, their recognition memory ratings fit the expected pattern (better recognition of Old items than New items). The Con18 - DEF participants, on the other hand, do
not appear to discriminate between Old and New, and Con18 – XYZ participants appear to favor New items over Old items. One possible reason for this is that they may have believed the New items X, Y, and Z to be more like the pseudorandomly structured sentences they saw during training. This speaks to a possible limitation in training controls with pseudorandom stimuli that is further described in Chapter 5.

Figure 4.5. Mean recognition memory scores for Old and New Items in the Exp18 and Con18 groups separated by New item type (DEF and XYZ). Lower scores represent stronger recognition ratings.

In order to assess whether the recognition memory test was contaminated by non-recognition memory based higher-order explicit knowledge or implicit knowledge (priming), two analyses were conducted: The first analysis looked at whether recognition memory performance was influenced in a way predicted by participants’ verbalizable microrules. The second
investigated whether recognition memory was present in the absence of priming. I describe both of these analyses in turn.

In order to assess the possible influence of higher-order explicit knowledge on the recognition memory test, comparisons were made on performance on Old and New items between verb final and non-verb final structures. That is, because so many participants reported a verb final (VF) microrule, the best way to ascertain an influence of microrules on the recognition memory task was to compare performance on VF and non-VF structures for the participants who reported using the VF microrule. If participants relied on a VF microrule, then they should differentially accept or reject items on the basis of their VF status (+/- VF), which would lead to correctly accepting +VF Old items (structures A [TSPOV] and B [TSOPV]) and correctly rejecting non-VF New items (Y-F [TVOSP], Z-D [STVPO], Z-F [TVPSO]). It would also lead to incorrect rejection of -VF Old items (structure C [TVSPO]) and incorrect acceptance of +VF New items (Y-D [TPOSV], Y-E [PTSOV], and Z-E [TPSOV]). Analyses were restricted to participants only reporting VF (n = 8) to avoid the influence of any other microrule knowledge (e.g., T-initial).

Table 4.6. Mean recognition memory ratings (SD in parentheses) for Old and New items by verb finality (+/- VF) for Exp18 participants who only reported the VF microrule.

<table>
<thead>
<tr>
<th></th>
<th>+VF</th>
<th>-VF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>1.62 (1.21)</td>
<td>3.37 (1.76)</td>
</tr>
<tr>
<td>New</td>
<td>3.38 (2.06)</td>
<td>3.70 (1.16)</td>
</tr>
</tbody>
</table>

A repeated measures ANOVA on recognition memory ratings was performed with Test Item Type (2 levels: Old and New), and Verb Finality (2 levels: +VF, -VF) as within-subjects factors. If participants primarily used the VF microrule to make their recognition memory judgments,
then there should be a significant effect of Verb Finality, but not of Test Item Type. If, however, participants did not exclusively use the VF microrule, then there should be a significant effect of Test Item Type. The repeated measures ANOVA revealed no effects of Verb Finality, $F(1, 7) = 2.58, p = .15$, and no interaction, $F(1, 7) = 1.59, p = .24$, but a significant effect of Test Item Type, $F(1, 7) = 5.15, p = .05, \eta^2_p = .42$. Thus, participants did not classify items exclusively on the basis of verb finality, but rather classified items primarily on the basis of their Old-New status. This result suggests that performance the recognition memory task cannot have been due exclusively to higher-order explicit knowledge as measured by the verbal reports.

To examine whether or not implicit knowledge in the form of priming influenced the results of the recognition memory task for the Exp18 group in the present experiment, mean reading times for whole sentences in the recognition memory test were compared. This was followed by a second, more important, comparison of reading times for test items that were classified by Exp18 participants as Old and New, regardless of their actual Old/New status. If participants were faster at reading test sentences that they subsequently classified as Old than test items classified as New, then they may have been using implicit priming to discriminate Old from New items (e.g., Shanks & Johnstone, 1999). If no such priming effect is present, though, then it suggests that implicit knowledge did not influence recognition memory performance, at least inasmuch as it stems from more efficient processing.

First, general evidence for the influence of implicit knowledge in the form of priming was assessed by comparing each Exp18 participant’s mean reaction time for Old and New sentences (see Table 4.7 below). Reading times were slower for actual Old items than actual New items, but the difference was not significant, $t(14) = 0.62, p = .54$. This rules out that priming occurred on actual Old and New test items. However, it was also possible that implicit
priming may have influenced when a participant judged test items to be Old or New, regardless of whether or not that item was truly Old or New. That is, participants may have implicitly used a priming-based fluency heuristic to make their recognition memory judgment. To assess this possibility, participants’ mean reaction time was computed for test sentences that they judged to be Old and judged to be New, regardless of the actual Old/New status of those sentences. Since participants did not uniformly use the recognition memory scale, their recognition memory judgments had to be converted to binary Old/New scores. To do this, each participant’s recognition memory scores were converted to z-scores. This procedure assumes that participants have an internal Old/New decision threshold (this is consistent with signal detection theory models of recognition memory, e.g., Berry et al., 2008). The z-score transformation converted this threshold to a value of zero, and participants’ recognition memory scores became positive and negative values around this zero threshold. In this scoring system, negative scores reflect Old judgments and positive scores reflect New judgments.

Table 4.7. Descriptive statistics for sentence reading times (in milliseconds) in the recognition memory task.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Old Items</td>
<td>12075</td>
<td>3766</td>
<td>972</td>
</tr>
<tr>
<td>Actual New Items</td>
<td>11701</td>
<td>2912</td>
<td>751</td>
</tr>
<tr>
<td>Items Judged Old</td>
<td>12214</td>
<td>3031</td>
<td>782</td>
</tr>
<tr>
<td>Items Judged New</td>
<td>11239</td>
<td>2924</td>
<td>755</td>
</tr>
</tbody>
</table>

36 This procedure assumes that the threshold metaphor for familiarity-based recognition memory is correct (Berry, Shanks, & Henson, 2008; Scott & Dienes, 2008; Shanks & Berry, 2012). Because this assumption may not be accurate, correlations were also performed between participants’ reading times and their recognition memory ratings. In the Exp18 group, there was a significant negative correlation between reading times and recognition memory ratings, \( r(90) = -0.20, p = .05 \). In other words, when reaction time decreased, recognition memory ratings went towards the New side of the scale. Thus, the correlation was in the opposite direction expected by an implicit priming account.
A paired-samples t-test was run on sentence reading times for items judged Old and items judged New. Reading times were not significantly different for items that participants rated as being Old or New, $t(14) = 1.38, p = .19$, 95% CI [-535, 2485]. Thus, there was no evidence of an influence of implicit priming or perceptual-motor fluency on the recognition memory task. Thus, participants do not appear to have used implicit priming as the basis for their recognition memory judgments.

The results of the recognition memory test in the 18-exposure condition support the hypothesis that participants would develop conscious knowledge of L2 syntax in the form of declarative memories after 18-exposures to the semiartificial language (six exposures to each syntactic structure). Difference score analyses revealed that Exp18 participants were able to discriminate Old and New items, while the Control group did not perform differently on either type of item. Moreover, there did not seem to be any robust contamination of the recognition memory results from either microrule knowledge or implicit priming, which indicates that genuine declarative memories supported performance on the recognition memory test.

### 4.3.1.3. Subjective fluency ratings

In order to assess whether increased exposure to training stimuli changed participants’ perception of the semiartificial language (i.e., to see if they felt like sentences were increasingly easier to read), fluency ratings were compared between and within subjects. A change in subjective fluency across the exposure phase would be taken as evidence for a change in participants’ lower-order explicit knowledge; however, it was predicted that after 18 exposures, the Exp18 participants would not develop lower-order conscious knowledge in the form of a change in subjective fluency. In order to obtain a measure of change in subjective fluency, difference
scores for each participant were computed as the difference between mean fluency judgments from Blocks 1 and 2 (see Table 4.6 below). These subjective fluency difference scores were normally distributed in both the Experimental group, $D(15) = 0.14, p = .20$, and the Control group, $D(14) = 0.11, p = .20$. An independent samples t-test revealed that fluency judgment difference scores were not significantly different in Experimental and Control groups, $t(27) = 0.98, p = .34$.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>Difference</td>
<td>-0.33</td>
<td>0.48</td>
<td>0.12</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.94</td>
<td>0.52</td>
<td>0.14</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.27</td>
<td>0.77</td>
<td>0.20</td>
<td>1.84</td>
</tr>
<tr>
<td>Control</td>
<td>Difference</td>
<td>-0.14</td>
<td>0.56</td>
<td>0.15</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.13</td>
<td>0.93</td>
<td>0.25</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.27</td>
<td>1.11</td>
<td>0.29</td>
<td>1.63</td>
</tr>
</tbody>
</table>

To determine whether the Exp18 or Con18 groups patterned differently in their subjective fluency ratings, a mixed ANOVA on subjective fluency ratings with Group as a between-subjects factor (2 levels: Experimental, Control) and Block as a within-subjects factor (2 levels: Block 1 and 2) revealed a main effect of Block, $F(1, 27) = 5.93, p = .02, \eta_p^2 = .18$, but no Group*Block interaction, $F(1, 27) = 0.95, p = .34$, and no effect of group, $F(1, 27) = 0.11, p = .74$. Post-hoc comparisons revealed no significant differences between groups on Block 1 or Block 2, $p > .30$. In terms of overall subjective fluency ratings, the Exp18 and Con18 groups did not differ, suggesting no learning-based changes in subjective fluency. However, the experimental group rated themselves as significantly less fluent on sentences in Block 2 than sentences in Block 1, $t(14) = 2.64, p = .02, d = 1.41$ (possible reasons for this are discussed in
Chapter 5). No other comparisons were significant. In general, in terms of subjective ease or difficulty of reading, participants reported that sentences either stayed the same throughout the exposure phase or got more difficult.

Figure 4.8. Mean fluency ratings in the Experimental and Control groups for each trial during the exposure phase. Lower numbers reflect a greater degree of ease in subjective fluency, while higher numbers reflect a greater degree of difficulty in subjective fluency.

Hypothesis 1 predicted that Exp18 participants would acquire conscious knowledge of L2 syntax in the form of higher-order explicit knowledge and declarative memories, but not by an increase in subjective ease-of-reading. The combined results of the retrospective verbal reports, recognition memory test, and subjective fluency ratings support hypothesis 1. However, there are open questions (e.g., about how to interpret an increase in subjective difficulty of reading) and
limitations (e.g., sample size issues), and these demand that the generalizability of the results be treated with caution.

4.3.2. Research question #2

In response to research question #2, hypotheses 2, 3, and 4 predicted that Experimental participants would acquire conscious knowledge of L2 syntax in the form of higher-order explicit knowledge and declarative memories in the 18-, 48-, and 96-exposure conditions and in the form of changes in subjective fluency in the 48- and 96-exposure conditions. The results reported in the previous section already demonstrated that participants formed conscious knowledge in the form of higher-order explicit knowledge and declarative memories, but not changes subjective fluency after just 18-exposures, which partially confirms hypotheses 2, 3, and 4. Therefore, this section focuses on the results of the 48- and 96-exposure conditions. Results are reported by focusing first on the verbal reports, followed by the recognition memory tests, and then the subjective fluency ratings.

4.3.2.1. Retrospective verbal reports

In order to assess whether participants acquired higher-order explicit knowledge in the 48- and 96-exposure phases, the retrospective verbal report interviews from the Exp48, Con48, Exp96, and Con96 groups were analyzed. Hypothesis 2 predicted that Exp48 and Exp96 participants would both demonstrate evidence of higher-order explicit knowledge (while Con48 and Con96 participants should not show any evidence of higher-order explicit knowledge). The results for the 48-exposure condition are reported first, followed by the 96-exposure condition.
Analysis of retrospective verbal report data from the Con48 group showed that 0 out of 13 participants verbalized any knowledge of the three target syntactic structures. In the Exp48, 8 out of 14 participants were able to verbally report some knowledge of the syntactic regularities of the three target syntactic structures. Examples of typical responses are shown in (5) and (6). Of the eight Exp48 participants who verbalized some syntactic knowledge, six of them mentioned the verb final regularity and three mentioned the initial Temporal Phrase regularity (one participant mentioned both). Thus, there was some evidence for the development of higher-order explicit knowledge of the three target syntactic structures in the Exp48, consistent with hypotheses 2.

(5) *I remembered some of them, because, like, certain words were at the end. Um, which one was, like the police car stopped, or something. Certain ones I was like, “oh, the verb is at the end,” but like since I had seen so many of the same sentence scrambled different ways, I really couldn’t tell for most of them if I had seen it or not.*

(6) *Sometimes the objects were at the end. Sometimes the verbs were at the end...like, the verb was at the end, so it didn’t confuse me too much.*

In the Con96 group, 0 out of 13 participants verbalized any knowledge of the three target syntactic structures. The analysis of the data from the Exp96 group showed that 13 out of 14 participants were able to verbally report some knowledge of the syntactic regularities of the three target syntactic structures. In the Exp96 group, 11 out of 13 participants reported a VF microrule (see 7 below for a typical example), and 9 of these 11 participants reported only the VF
microrule, while 2 of these 11 also reported awareness that sentences could begin with the TS bigram (see 8 below for an example).

(7) Later on, I was thinking that it seemed like if the verb was at the end, it was even harder [to comprehend]…At the end it seemed like it was doing the the verb was at the end more.

(8) It seemed more often than not the main noun, what was performing the action, was close to the beginning, like not all the time, like, maybe 60% of the time [interviewer: the subject?] yeah, the subject, yeah, after “some time ago”…time words, at the beginning most of the time

As in the Exp18 and Exp48 groups, there was some evidence for the development of higher-order explicit knowledge of the three target syntactic structures in the Exp96. Taken together, these findings are consistent with hypothesis 2: participants developed higher-order explicit knowledge regardless of the amount of input during the exposure phase.

4.3.2.2. Recognition memory tests

In order to assess whether participants acquired conscious knowledge in the form of declarative memories in the 48- and 96-exposure phases, the recognition memory tests from the Exp48, Con48, Exp96, and Con96 groups were analyzed. Hypothesis 3 predicted that Exp48 and Exp96 participants would both demonstrate evidence of declarative memory for both Late items (sentences presented in the last 18 trials of the exposure phase) and Early items (sentences presented in the first 18 trials of the exposure phase). The results for the 48-exposure condition are reported first, followed by the 96-exposure condition. As with the 18-exposure condition...
reported above, a difference score analysis precedes the omnibus analyses on participants’ mean recognition ratings. These analyses are then followed with analyses that check for contamination of the recognition memory scores by higher-order explicit knowledge and implicit priming.

As with the 18-exposure condition, recognition memory difference scores were computed for each participant by subtracting each participant’s mean rating for New test items minus his/her mean rating for Old test items. Each participant was given two difference scores, an early difference score and a late difference score. These data are reported in Tables 4.9 and 4.10 below. The difference between the two types of score was simply whether the core sentences for those test items came from the initial phase of training (Early) or the final phase (Late). This variable was labeled “Source.”

Table 4.9. Descriptive statistics for recognition memory difference scores by group in the E48 condition.

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp48</td>
<td>Early</td>
<td>0.85</td>
<td>4.00</td>
<td>1.09</td>
<td>0.29</td>
<td></td>
<td>0.22</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>1.78</td>
<td>2.67</td>
<td>0.82</td>
<td>0.21</td>
<td></td>
<td>1.31</td>
<td>2.26</td>
</tr>
<tr>
<td>Con48</td>
<td>Early</td>
<td>0.25</td>
<td>5.33</td>
<td>1.67</td>
<td>0.46</td>
<td></td>
<td>-0.75</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>-0.16</td>
<td>4.33</td>
<td>1.18</td>
<td>0.32</td>
<td></td>
<td>-0.88</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Table 4.10. Results of tests to check for the normality of distributions of data on the recognition memory difference scores in the E48 condition.

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Kolmogorov-Smirnov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Skew</td>
<td>SE</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Exp48</td>
<td>Early</td>
<td>0.91</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>0.57</td>
<td>0.59</td>
<td>-1.01</td>
</tr>
<tr>
<td>Con48</td>
<td>Early</td>
<td>0.37</td>
<td>0.62</td>
<td>-0.67</td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>1.14</td>
<td>0.62</td>
<td>1.83</td>
</tr>
</tbody>
</table>

In order to assess how participants in the 48-exposure condition classified recognition memory test items, a mixed ANOVA with Group (2 levels: Experimental, Control) as a between-subjects variable and Source (2 levels: Early, Late) as within-subjects variable was run with difference scores as the dependent variable. The ANOVA revealed no effect of Source, $F(1, 25) = 0.84, p = .37$, suggesting that the Late-Early status of test items did not contribute to differences in performance in the same way in the Exp48 and Con48 groups. The ANOVA also revealed a significant effect of Group, $F(1, 25) = 11.16, p = .003, \eta^2_p = .31$, and a significant Group*Source interaction, $F(1, 25) = 5.97, p = .02, \eta^2_p = .19$, suggesting that the Exp48 and Con48 groups patterned differently in their performance on Early and Late items (Figure 4.7). To further investigate this interaction effect, post-hoc comparisons were conducted to determine whether difference scores were significantly different between groups. The Exp48 group had significantly larger difference scores for Late items than the Con48 group, $U = 20.00, z = -3.45, p < .001, r = .66$. The Exp48 group also had larger difference scores for Early items than the Con48 group, but this difference was not significant, $U = 64.50, z = -1.29, p = .20$. By this measure, recognition memory in the Exp48 was only better than the Con48 for Late items.
Once it was established that Exp48 participants had recognition memory, it was next important to determine whether there were between-group differences in recognition memory for specific structures. To that end, the next analyses investigated whether Exp48 and Con48 participants differed with respect to how they rated the individual syntactic structures in the recognition memory task (Table 4.11 below show the descriptive statistics for these data). The structures D, E, and F were used for Early-New items, while the New structures X, Y, and Z were used for Late-New items. Because the New structures were not identical, two separate mixed ANOVAs were carried out, one on Early items and one on Late items.

A 2x6 mixed ANOVA on recognition memory ratings for Early items with Group (Experimental, Control) as between-subjects factor and Structure (A, B, C, D, E, F) as within-
subjects factors revealed a significant effect of Structure, \(F(5, 125) = 4.18, p = .002, \eta_p^2 = .14\), but no significant effect of Group, \(F(1, 25) = 0.31, p = .58\), and no Structure*Group interaction, \(F(5, 125) = 0.82, p = .53\). Closer inspection of the descriptive statistics shows that for Early items, structure F (TVOSP) was rated as least recognized by both the Exp48 participants and the Con48 participants. Bonferroni-adjusted post-hoc comparisons investigating this difference support this conclusion. In the Exp48 group, structure F was rated significantly less recognized than structure A, B, and D. No other comparisons were significant. Likewise, no comparisons between structures within the Con48 group were significant, \(ps > .11\). Overall, the findings suggest no robust recognition memory for Early items in the Exp48 group over the Con48 group, consistent with the results of the recognition memory difference score analysis.

A 2x6 mixed ANOVA on recognition memory ratings for Late items with Group (Experimental, Control) as between-subjects factor and Structure (A, B, C, X, Y, Z) as within-subjects factors revealed a significant effect of Structure, \(F(5, 115) = 5.81, p < .001, \eta_p^2 = .20\), and a significant Group*Structure interaction, \(F(5, 115) = 4.94, p < .001, \eta_p^2 = .18\), but no effect of Group, \(F(1, 23) = 1.69, p = .20\). The Structure and Group*Structure interaction effects indicate that the Exp48 and Con48 groups patterned differently in their recognition memory for Late items, again consistent with the recognition memory difference score analyses. Post-hoc comparisons revealed significant differences between groups on Late items for structure C, \(t(25) = 2.37, p = .02, d = .94\), structure X, \(t(25) = 2.19, p = .04, d = .87\), and structure Y, \(t(25) = 2.53, p = .02, d = 1.01\). No other significant differences in mean endorsements were found between groups on Late items.
Table 4.11. Descriptive statistics for the recognition memory task in the 48 exposure condition by group and test item type. K-S tests revealed the data reported below to be normally distributed, all $p$s > .09.

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Test Item Type</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Early</td>
<td>Old</td>
<td>2.59</td>
<td>2.33</td>
<td>0.63</td>
<td>0.17</td>
<td>2.23</td>
<td>2.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>2.42</td>
<td>4.00</td>
<td>1.22</td>
<td>0.32</td>
<td>1.72</td>
<td>3.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.57</td>
<td>3.00</td>
<td>0.85</td>
<td>0.22</td>
<td>2.07</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2.78</td>
<td>3.00</td>
<td>1.05</td>
<td>0.28</td>
<td>2.17</td>
<td>3.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>3.45</td>
<td>2.33</td>
<td>0.78</td>
<td>0.21</td>
<td>3.00</td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>2.57</td>
<td>4.00</td>
<td>1.39</td>
<td>0.37</td>
<td>1.76</td>
<td>3.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>3.28</td>
<td>5.00</td>
<td>1.54</td>
<td>0.41</td>
<td>2.39</td>
<td>4.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>4.50</td>
<td>5.00</td>
<td>1.60</td>
<td>0.42</td>
<td>3.57</td>
<td>5.42</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Early</td>
<td>Old</td>
<td>3.00</td>
<td>2.33</td>
<td>0.80</td>
<td>0.22</td>
<td>2.51</td>
<td>3.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>3.00</td>
<td>4.00</td>
<td>1.08</td>
<td>0.29</td>
<td>2.34</td>
<td>3.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.92</td>
<td>5.00</td>
<td>1.32</td>
<td>0.36</td>
<td>2.12</td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>3.07</td>
<td>5.00</td>
<td>1.75</td>
<td>0.48</td>
<td>2.01</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>3.25</td>
<td>3.67</td>
<td>1.14</td>
<td>0.31</td>
<td>2.56</td>
<td>3.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>3.00</td>
<td>4.00</td>
<td>1.41</td>
<td>0.39</td>
<td>2.14</td>
<td>3.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>2.92</td>
<td>5.00</td>
<td>1.80</td>
<td>0.49</td>
<td>1.83</td>
<td>4.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>3.84</td>
<td>3.00</td>
<td>1.21</td>
<td>0.33</td>
<td>3.11</td>
<td>4.57</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>Late</td>
<td>Old</td>
<td>2.00</td>
<td>2.00</td>
<td>0.74</td>
<td>0.19</td>
<td>1.57</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>2.00</td>
<td>2.00</td>
<td>0.73</td>
<td>0.21</td>
<td>1.53</td>
<td>2.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.42</td>
<td>4.00</td>
<td>1.50</td>
<td>0.40</td>
<td>1.55</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>1.27</td>
<td>3.00</td>
<td>0.93</td>
<td>0.25</td>
<td>1.03</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>3.78</td>
<td>3.33</td>
<td>0.91</td>
<td>0.24</td>
<td>3.25</td>
<td>4.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>3.28</td>
<td>5.00</td>
<td>1.85</td>
<td>0.49</td>
<td>2.21</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y</td>
<td>4.21</td>
<td>4.00</td>
<td>1.57</td>
<td>0.42</td>
<td>3.30</td>
<td>5.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z</td>
<td>3.85</td>
<td>4.00</td>
<td>1.29</td>
<td>0.34</td>
<td>3.11</td>
<td>4.60</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Late</td>
<td>Old</td>
<td>2.70</td>
<td>1.67</td>
<td>0.57</td>
<td>0.16</td>
<td>2.35</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>1.92</td>
<td>3.00</td>
<td>0.95</td>
<td>0.26</td>
<td>1.34</td>
<td>2.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.84</td>
<td>5.00</td>
<td>1.67</td>
<td>0.46</td>
<td>1.83</td>
<td>3.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2.61</td>
<td>4.00</td>
<td>1.32</td>
<td>0.37</td>
<td>1.81</td>
<td>3.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>2.53</td>
<td>4.00</td>
<td>1.14</td>
<td>0.31</td>
<td>1.84</td>
<td>3.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>1.94</td>
<td>6.00</td>
<td>1.51</td>
<td>0.42</td>
<td>0.92</td>
<td>2.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y</td>
<td>2.53</td>
<td>6.00</td>
<td>1.85</td>
<td>0.51</td>
<td>1.41</td>
<td>3.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z</td>
<td>3.23</td>
<td>5.00</td>
<td>1.48</td>
<td>0.41</td>
<td>2.33</td>
<td>4.12</td>
<td></td>
</tr>
</tbody>
</table>

Overall, these findings mirror that of the between-group difference score comparisons:

the Exp48 group differed from the Con48 group on Late items, but not Early items. Inspection of
Figure 4.8 below appears to show that Exp48 group discriminates Old and New Early items; however, this difference is not robust enough to reach significance.

![Figure 4.8](image)

**Figure 4.8.** Mean recognition memory ratings for Old and New items by Source (Early, Late) in the Exp48 and Con48 groups. Lower scores represent strong recognition ratings.

In order to assess whether the recognition memory test for the Exp48 group was contaminated by higher-order explicit knowledge or implicit knowledge (priming), two analyses were conducted: The first looked at whether recognition memory performance went beyond that which can be accounted for by the retrospective verbal report data. The second investigated
whether recognition memory was present in the absence of priming. I describe both of these analyses in turn.

To assess whether or not Exp48 participants might have used verbalizable knowledge on the recognition memory test, Old-New difference scores were analyzed with respect to whether participants were aware or unaware in their verbal reports (an analysis that was not possible in the 18-exposure condition, since 14/15 participants were aware). In this analysis, Aware participants (n = 8) were those who verbalized any knowledge of the three target syntactic structures. Participants who verbalized no knowledge consistent with the three target syntactic structures were classified as Non-Verbalizing (n = 6). These participants were not classified as Unaware, since they may well have been aware by other measures of awareness. If participants have recognition memory beyond that found in the verbal reports, then Non-Verbalizing participants should be able to perform above the Con48 group.

Non-Verbalizing Exp48 participants were compared with Con48 participants to determine whether they could perform better than Controls. A mixed ANOVA on difference scores with Group (Non-Verbalizing, Con48) as between-subjects factor and Source (Late, Early) as within-subjects factor revealed a significant effect of Group, $F(1, 17) = 4.44, p = .05, \eta_p^2 = .21$, but no significant effect of Source and no Source*Group interaction, $Fs < 1$. Post-hoc comparisons indicated that difference scores on Late items for the Non-Verbalizing Exp48 participants ($M_{\text{Late}} = 1.61, SD = 1.02$) were significantly larger than those in the Control group ($M_{\text{Late}} = -0.16, SD = 1.18$), $t(17) = 3.16, p = .006, d = 1.53$; however, difference scores on Early items for the Non-Verbalizing Exp48 participants ($M_{\text{Early}} = 0.83, SD = 1.37$) were not significantly larger than those in the Control group ($M_{\text{Early}} = 0.25, SD = 1.67$), $t(17) = .73, p = .47$. These results indicate that participants who did not express any awareness of the three
syntactic structures in the retrospective verbal reports were still able to perform better than the Con48 group on the recognition memory test, but only on Late items (see Figure 4.9 below). Thus, there was evidence for recognition memory beyond what would be predicted if learners relied exclusively on the higher-order explicit knowledge expressed in the verbal reports.

Figure 4.9. Mean difference scores on Late and Early items in the 48-exposure condition across Aware, Non-Verbalizing, and Control participants.

Did the Aware participants perform any better than the Con48 group on Late and Early items? Difference scores on Late items for the Aware Exp48 participants ($M_{Late} = 1.91$, $SD = 0.68$) were significantly larger than those in the Control group, $t(19) = 2.64$, $p = .01$, $d = 1.21$; however, difference scores on Early items for the Aware Exp48 participants ($M_{Early} = 0.87$, $SD = 0.92$) were not significantly larger than those in the Control group, $t(19) = .95$, $p = .35$. Taken
together, the results indicate that Exp48 participants had recognition memory for Late items, but not Early items, regardless of whether they expressed awareness in their verbal reports. Although picture is made more complex by the small sample size, it nevertheless is consistent with the interpretation that recognition memory data reflect memory-based knowledge beyond the use of higher-order explicit knowledge.

To assess whether there was a priming effect on Old and New items in the recognition memory test, each Exp48 participant’s mean reading times for sentences was computed separately for Late Old and New items and Early Old and New items. If participants were significantly faster at reading Old vs. New sentences on the recognition memory test, then it would suggest that priming may have been involved in their performance. Reading times were faster for Late Old items ($M = 13947, SD = 2763$) than Late New items ($M = 14295, SD = 3169$), but the difference was not significant, $t(13) = .71$, 95% CI [-2314, 1616]. Reading times for Early Old items ($M = 13839, SD = 3373$) were slower than Early New items ($M = 13115, SD = 3114$), but the difference was not significant, $t(13) = 1.80$, $p = .09$, 95% CI [-142, 1591]. Thus, there was not a significant priming effect on participants’ ratings of Old and New items in the recognition memory test.

In order to assess whether participants classified test items as Old or New on the basis of implicit priming (regardless of the actual Old-New status of the test items), participants’ mean RTs for Late items they judged to be Old and New and Early items they judged to be Old and New were submitted to paired samples t-tests. If participants judge Old vs. New items on the basis of an implicit priming heuristic, then they should be faster at reading sentences they judge

---

As noted earlier, some may object to the z-transformation procedure to assess priming effects on recognition memory. To account for such objections, correlations were also run to assess any effect of priming on recognition memory judgments. Correlations between RT and recognition memory ratings were run with Spearman’s rho, and the results showed no significant correlation between RT and recognition memory ratings, $\rho(270) = .06$, $p = .15$. 

157
to be old, regardless of whether those items were actually Old or not. The results show that
Exp48 participants were faster at reading Late items judged Old ($M = 13957, SD = 2384$) than
Late items judged New ($M = 14566, SD = 4665$), but this difference was not significant, $t(13) = .47, p = .64, 95\% \text{ CI } [-3409, 2189]$. Exp48 participants also were not significantly faster at
reading Early items judged Old ($M = 13408, SD = 3313$) than Early items judged New ($M = 13611, SD = 3302$), $t(13) = .42, p = .68, 95\% \text{ CI } [-1244, 839]$. Thus, there was no evidence of
the influence of implicit knowledge in the form of priming.

The results suggest that the Exp48 participants developed conscious knowledge in the
form of declarative memories but only for Late items. Subsequent analyses revealed that this
effect could not be accounted for exclusively by higher-order explicit knowledge or implicit
priming. Therefore, participants in the Exp48 group appeared to use genuine declarative
memories as the basis of their recognition memory decisions.

Analyses of the recognition memory test performance in the 96-exposure condition
followed the same rationale and protocol as the 48-exposure condition. Difference scores were
computed for each participant by subtracting each participant’s mean rating for New test items
minus his/her rating for Old test items. Each participant was given two difference scores, an
early difference score and a late difference score. These data are reported in Table 4.12 below.
The difference between the two types of score was simply whether the core sentences for those
test items came from the 18 trials of the exposure phase (Early) or the final 18 trials of the
exposure phase (Late). This variable was labeled “Source.”
Table 4.12. Descriptive statistics for difference scores by group in the E96 condition.

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp96</td>
<td>Early</td>
<td>1.54</td>
<td>6.33</td>
<td>1.60</td>
<td>0.41</td>
<td>0.64</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>1.05</td>
<td>3.33</td>
<td>0.91</td>
<td>0.23</td>
<td>0.54</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>Con96</td>
<td>Early</td>
<td>0.46</td>
<td>4.33</td>
<td>1.28</td>
<td>0.34</td>
<td>-0.27</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>0.30</td>
<td>3.33</td>
<td>0.98</td>
<td>0.26</td>
<td>-0.26</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

K-S tests showed that the difference scores in both the Experimental group and Control group were normally distributed, all $p$s > .08. In order to assess the difference scores for Late and Early items across groups, an omnibus mixed ANOVA with Group (2 levels: Exp96, Con96) as between-subjects factor and Source (2 levels: Early, Late) as within-subjects factor was performed on difference scores. The ANOVA revealed a significant effect of Group, $F(1, 25) = 6.41$, $p = .01$, $\eta^2_p = .21$, but no effect of Source, $F(1, 25) = 0.89$, $p = .35$, and no Source*Group interaction, $F(1, 25) = 0.22$, $p = .64$. The mean difference scores on Early and Late items were larger for Exp96 participants than Con96 participants, indicating the presence of recognition memory in the Exp96 group (Figure 4.10 below).

To further investigate the between-groups differences in recognition memory difference scores, post-hoc comparisons were performed using independent-samples t-tests. Results showed significantly larger difference scores for the Exp96 group over the Con96 group on both Late items, $t(25) = 2.13$, $p = .04$, $d = .85$, and Early items, $t(25) = 2.00$, $p = .05$, $d = .76$. Thus, the Exp96 had recognition memory for both Early and Late items, contrary to the Exp48 group, which only had recognition memory for Late items.\(^{38}\)

Finally, within-subjects analyses on

\(^{38}\) Comparable analyses using mean recognition memory ratings, instead of difference scores, in the Exp96 group revealed significant differences in recognition memory for Late Old vs. New items, $t(13) = 4.16$, $p = .001$, $d = 1.17$, and Early Old vs. New items, $t(13) = 3.44$, $p = .004$, $d = 1.28$. There were no differences in recognition memory ratings for Old vs. New sentences for either item type in the Con96 group, $p$s > .23.
difference scores using paired-sample t-tests showed no difference in performance on Late and Early items in the Experimental group, $t(13) = 1.00, p = .33$, or the Control group, $t(12) = .33, p = .74$, which indicates that neither group performed better or worse on Late vs. Early items.

Figure 4.10. Mean difference scores for Early and Late items in the recognition memory task for the Experimental and Control groups in the Exp96 condition. Larger scores reflect more accurate discrimination of Old and New items.

Once it was established that Exp96 participants had recognition memory, it was next important to determine whether there were between-group differences in recognition memory for specific structures in Late and Early items. To that end, the next analyses investigated whether Exp96 and Con96 participants differed with respect to how they rated the individual syntactic structures in the recognition memory task (Table 4.13 below shows the descriptive statistics for these data). A 2x6 mixed ANOVA on mean recognition memory ratings (as opposed to
difference scores) on Early items with Group (2 levels: Exp96, Con96) as between-subjects variable and Structure (6 levels: A, B, C, D, E, F) as within-subjects variable revealed a significant effect of Structure, $F(5, 115) = 4.84, p < .001, \eta_p^2 = .17$, and no significant effect of Group, $F(1, 23) = 0.45, p = .51$, and no Group*Structure interaction, $F(5, 115) = 1.58, p = .17$. Bonferroni-adjusted post-hoc comparisons on the Exp96 group showed significant differences in recognition memory ratings between structures B and F, $p = .001$, and C and F, $p = .03$, and no significant differences in recognition memory ratings between any structures in the Con96 group. Thus, the effect of Structure on mean recognition memory ratings appears to be based on the Exp96 group’s ability to accurately rate structures B and C as Old and structure F as New.

Figure 4.11. Mean recognition memory ratings for Old and New items by Source (Early, Late) in the Experimental and Control groups in the Exp96 condition. Lower scores represent stronger recognition ratings.
Table 4.13. Descriptive statistics for the recognition memory task in the Exp96 condition by group and test item type. K-S tests reveal the data to be normally distributed.

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Test Item Type</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Experimental</td>
<td>Early</td>
<td>Old</td>
<td>2.09</td>
<td>4.00</td>
<td>0.99</td>
<td>0.25</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>2.85</td>
<td>5.00</td>
<td>2.10</td>
<td>0.56</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>1.50</td>
<td>2.00</td>
<td>0.75</td>
<td>0.20</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>1.92</td>
<td>5.00</td>
<td>1.43</td>
<td>0.38</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>3.63</td>
<td>4.00</td>
<td>1.30</td>
<td>0.33</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>3.21</td>
<td>5.00</td>
<td>1.71</td>
<td>0.46</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>3.28</td>
<td>5.00</td>
<td>1.63</td>
<td>0.43</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>4.46</td>
<td>5.00</td>
<td>1.80</td>
<td>0.50</td>
<td>3.36</td>
</tr>
<tr>
<td>Control</td>
<td>Early</td>
<td>Old</td>
<td>2.75</td>
<td>4.00</td>
<td>1.17</td>
<td>0.31</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>3.16</td>
<td>5.00</td>
<td>1.64</td>
<td>0.44</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.77</td>
<td>5.00</td>
<td>1.84</td>
<td>0.49</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2.62</td>
<td>5.00</td>
<td>1.90</td>
<td>0.51</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>3.22</td>
<td>3.51</td>
<td>0.84</td>
<td>0.22</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>3.51</td>
<td>5.00</td>
<td>1.64</td>
<td>0.46</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>3.16</td>
<td>5.00</td>
<td>1.41</td>
<td>0.37</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>3.38</td>
<td>5.00</td>
<td>1.21</td>
<td>0.32</td>
<td>2.68</td>
</tr>
<tr>
<td>Experimental</td>
<td>Late</td>
<td>Old</td>
<td>1.90</td>
<td>1.67</td>
<td>0.64</td>
<td>0.16</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>1.81</td>
<td>5.00</td>
<td>1.35</td>
<td>0.36</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>1.14</td>
<td>1.00</td>
<td>0.36</td>
<td>0.09</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2.71</td>
<td>5.00</td>
<td>1.77</td>
<td>0.47</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>2.96</td>
<td>3.17</td>
<td>1.04</td>
<td>0.27</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>2.67</td>
<td>4.00</td>
<td>1.37</td>
<td>0.39</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y</td>
<td>2.57</td>
<td>5.00</td>
<td>1.89</td>
<td>0.49</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z</td>
<td>3.50</td>
<td>5.00</td>
<td>1.95</td>
<td>0.52</td>
<td>2.37</td>
</tr>
<tr>
<td>Control</td>
<td>Late</td>
<td>Old</td>
<td>2.56</td>
<td>3.33</td>
<td>1.01</td>
<td>0.27</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>2.46</td>
<td>4.00</td>
<td>1.54</td>
<td>0.41</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>3.16</td>
<td>5.00</td>
<td>1.72</td>
<td>0.47</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2.89</td>
<td>4.00</td>
<td>1.38</td>
<td>0.39</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New</td>
<td>2.87</td>
<td>4.00</td>
<td>1.08</td>
<td>0.28</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>3.45</td>
<td>5.00</td>
<td>1.55</td>
<td>0.41</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y</td>
<td>3.60</td>
<td>5.00</td>
<td>1.49</td>
<td>0.39</td>
<td>2.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z</td>
<td>2.54</td>
<td>5.00</td>
<td>1.49</td>
<td>0.41</td>
<td>1.67</td>
</tr>
</tbody>
</table>

To investigate the effects of individual structures on Exp96 and Con96 participants on Late items, a 2x6 mixed ANOVA on mean recognition memory ratings (as opposed to difference
scores) on Late items with Group (2 levels: Exp96, Con96) as between-subjects variable and Structure (6 levels: A, B, C, X, Y, Z) as within-subjects variable was conducted. The ANOVA revealed a significant effect of Group*Structure interaction, $F(5, 100) = 2.42, p < .04, \eta^2_p = .11$, and no significant effect of Group, $F(1, 20) = 2.24, p = .15$, or Structure, $F(5, 100) = 1.66, p = .15$. Post-hoc comparisons investigated whether the Exp96 and Con96 groups differed in their ratings for specific Late structures. Exp96 participants rated Structure B as more recognized than Con96 participants, $t(24) = 4.12, p < .001, d = 1.68$; however, there were no other differences between groups on the other Late structures.

In order to assess whether the recognition memory test for the Ex96 group was contaminated by non-recognition memory based higher-order explicit knowledge or implicit knowledge (priming), two analyses were conducted: The first looked at whether recognition memory performance went beyond that which can be accounted for by the retrospective verbal report data. The second investigated whether recognition memory was present in the absence of priming. I describe both of these analyses in turn.

In the Exp96 group, 13 out of 14 participants reported higher-order explicit knowledge. 9 participants only reported higher-order explicit knowledge of a verb final microrule (VF). Because so many participants reported the VF regularity, the only way to ascertain an influence of higher-order explicit knowledge on the recognition memory test was to compare performance across VF and non-VF structures for the participants who reported using the VF microrule (as was done for the Exp18 group). If participants relied on a VF microrule, then they should differentially accept or reject items on the basis of their VF status (+/- VF), which would lead to correctly accepting +VF Old items (structures A [TSPOV] and B [TSOPV]) and correctly rejecting non-VF New items (Early: F [TVOSP]; Late X [STVPO], Z [TVPSO]). It would also
lead to incorrect rejection of -VF Old items (structure C [TVSPO]) and incorrect acceptance of +VF New items (Early: D [TPOSV], E [PTSOV]; Late: Y [TPSOV]). To avoid confounding microrule knowledge of the VF regularity with knowledge of other microrules, analyses were restricted to participants who only reported awareness of the VF regularity (n = 9). A 2x2x2 repeated measures ANOVA on recognition memory ratings with Source (2 levels: Late, Early), Test Item Type (2 levels: Old and New), and Verb Finality (2 levels: +VF, -VF) as within-subjects factors revealed no effects of Source or Verb Finality, ps > .21, and no interactions, ps > .58, but a significant effect of Test Item Type, $F(1, 8) = 7.06, p = .03, \eta^2_p = .47$. Participants discriminated Old v. New sentences on the basis of the Old-New status of the items, while higher-order explicit knowledge of the VF regularity did not clearly influence the Exp96 group’s performance on the recognition memory test. In other words, performance on the recognition memory task cannot have been due exclusively to higher-order explicit knowledge as measured by the verbal reports.

To assess whether there was a priming effect on Old and New items in the recognition memory test, each Exp96 participant’s mean reading times for sentences was computed separately for Late Old and New items and Early Old and New items. If participants were significantly faster at reading Old vs. New sentences, then it would suggest that priming may have been involved in their performance. Reading times were faster for Late Old items ($M = 15394, SD = 6957$) than Late New items ($M = 15913, SD = 4937$), but the difference was not significant, $t(13) = .35, p = .72, 95\% CI [-3655, 2618]$. Reading times for Early Old items ($M = 13664, SD = 4642$) was faster than Early New items ($M = 14189, SD = 4127$), but the difference was not significant, $t(13) = .56, p = .58, 95\% CI [-2541, 1490]$. Thus, there was not a significant priming effect on participants’ ratings of Old and New items in the recognition memory test.
In order to assess whether participants classified test items as Old or New on the basis of implicit priming (regardless of the actual Old-New status of the test items), participants’ mean RTs for Late items they judged to be Old and New and Early items they judged to be Old and New were submitted to paired samples t-tests. If participants judged Old vs. New items on the basis of an implicit priming heuristic, then they should be faster at reading items they judge to be old, regardless of whether those items were actually Old or not. Following the same z-transformation procedure as in the E18 and E48 analyses, participants’ mean RTs for Late and Early recognition memory test items they judged to be Old vs. New were submitted to a paired-samples t-test. Participants’ mean RTs for Late items judged Old (M = 15379, SD = 6052) was faster than Late items judged New (M = 16448, SD = 5816), but this difference was not significant, t(13) = 0.79, p = .44, 95% CI [-4001, 1892]. Mean RTs for Early items judged Old (M = 13984, SD = 4642) were slower than Early items judged New (M = 13879, SD = 4117), but this difference was not significant, t(13) = 0.15, p = .87, 95% CI [-1343, 1553]. Therefore, no evidence of a robust effect of implicit knowledge through priming was obtained.

The results suggest that the Exp96 participants developed conscious knowledge in the form of declarative memories for both Early and Late items. Unlike the E48 group, Exp96 participants were able to discriminate Old vs. New Early items better than Controls. Subsequent analyses revealed that this effect could not be accounted for exclusively by higher-order explicit knowledge or implicit priming. Therefore, participants in the Exp96 group appeared to use genuine declarative memories as the basis of their recognition memory decisions. Now, hypothesis 3 predicted that Experimental participants would acquire conscious knowledge in the form of declarative memory (as measured by the recognition memory test), regardless of the

39 In keeping with the earlier analyses, correlations were also run to assess any effect of priming on recognition memory judgments. Correlations between RT and recognition memory ratings were run with Spearman’s rho, and the results showed no significant correlation between RT and recognition memory ratings, $\rho(270) = .08, p = .14$. 

165
duration of their exposure phases. Taken together, the results from the Exp18, Exp48, and Exp96 groups demonstrate evidence of conscious knowledge in the form declarative memory for Late items (sentences presented in the last 18 trials of the exposure phase), and some evidence of such conscious knowledge for Early items (sentences presented in the first 18 trials of the exposure phase). Although the lack of statistically robust evidence that the Exp48 condition had recognition memory for Early items must be taken into consideration, the overall pattern of results support hypothesis 3.

4.3.2.3. Subjective fluency ratings

In order to assess whether participants acquired conscious knowledge in the form of trial-by-trial changes to subjective fluency, subjective fluency ratings were analyzed for the Exp48, Con48, Exp96, and Con96 groups. Hypothesis 4 predicted that Exp48 and Exp96 participants would acquire lower-order explicit knowledge as indicated by increases in subjective ease of reading. It was also predicted that, due to the subtle nature of changes in fluency (Perruchet, 2008), participants in the 18-exposure condition would not experience changes in subjective fluency. The detailed results of the subjective fluency ratings in the 18-exposure condition are reported in section 4.3.1.3 above. Overall, Exp18 participants were not significantly different from the Con18 participants in their subjective fluency ratings, although the Exp18 participants did show a small, but significant increase in perceived difficulty of reading the exposure phase sentences, contrary to Hypothesis 4. To further address hypothesis 4, the present analyses focused on the subjective fluency rating data from the 48- and 96-exposure conditions.

In order to assess whether or not increased exposure to training stimuli resulted in changes to participants lower-order explicit knowledge (i.e., to see if they felt like sentences were
increasingly easier to read), fluency ratings were compared between and within subjects.

Subjective fluency rating difference scores for each participant were computed as the difference between mean fluency judgments from the first 24 training trials and the last 24 training trials (see Table 4.14 below). The subjective fluency rating difference scores were normally distributed in both the Experimental group, $D(14) = 0.18, p = .20$, and the Control group, $D(13) = 0.10, p = .20$.

Table 4.14. Descriptive statistics for the subjective fluency ratings during the training phase in the 48 exposure condition by group and test item type.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bound</td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>2.34</td>
<td>2.44</td>
<td>0.60</td>
<td>0.16</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.49</td>
<td>2.22</td>
<td>0.60</td>
<td>0.16</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.69</td>
<td>2.33</td>
<td>0.73</td>
<td>0.19</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.71</td>
<td>2.78</td>
<td>0.78</td>
<td>0.21</td>
<td>2.25</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.28</td>
<td>1.67</td>
<td>0.51</td>
<td>0.13</td>
<td>-0.57</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>2.50</td>
<td>3.56</td>
<td>1.04</td>
<td>0.29</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.52</td>
<td>4.00</td>
<td>0.97</td>
<td>0.27</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.56</td>
<td>3.11</td>
<td>0.89</td>
<td>0.24</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.56</td>
<td>3.11</td>
<td>0.82</td>
<td>0.22</td>
<td>2.06</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.02</td>
<td>1.65</td>
<td>0.55</td>
<td>0.15</td>
<td>-0.35</td>
</tr>
</tbody>
</table>

An independent samples t-test revealed that subjective fluency ratings difference scores were not significantly different in Experimental and Control groups, $t(25) = 1.27, p = .21$. In order to determine whether the Exp48 and Con48 groups patterned differently in the subjective fluency ratings, analysis of the mean subjective fluency scores in each group (as opposed to difference scores) was conducted with a 2x4 mixed ANOVA with Group (Experimental, Control) as a
between-subjects factor and Block (Blocks 1-4) as a within-subjects factor. The ANOVA revealed no significant effects of Block, $F(2.06, 51.55) = 1.22, p = .30$, Group, $F(1, 25) = .001, p = .97$, and no Block*Group interaction, $F(2.06, 51.55) = 1.10, p = .35$. These results provide no evidence of a change in subjective fluency in the Exp48 group, contrary to the predictions of hypothesis 4 (see Figure 4.12 for an illustration).

![Figure 4.12](image.png)

Figure 4.12. Mean subjective fluency ratings in the Experimental and Control groups for each trial in the exposure phases of the E48 condition. Lower numbers reflect a greater degree of subjective ease of reading.

The same analyses were conducted on the subjective fluency ratings data from the Exp96 and Con96 groups (see Table 4.15 below). The difference scores (ratings for the first 24 trials minus ratings for the final 24 trials in the exposure phase) were normally distributed in both the Experimental group, $D(14) = 0.16, p = .20$, and the Control group, $D(13) = .21, p = .11$. An
independent samples t-test revealed that subjective fluency ratings difference scores were not significantly different in Experimental and Control groups, \( t(25) = 0.42, p = .67 \), indicating that there were no increases in subjective ease of reading in the Exp96 group. In order to determine whether the Exp96 and Con96 groups patterned differently in the subjective fluency ratings, an analysis of the mean subjective fluency scores in each group (as opposed to difference scores) was conducted with a 2x8 mixed ANOVA with Group (Experimental, Control) as a between-subjects factor and Block as a within-subjects factor (Blocks 1-8). The ANOVA with Greenhouse-Geisser correction revealed no significant effects of Block, \( F(4.21, 105.44) = .99, p = .43 \), Group, \( F(1, 25) = .001, p = .97 \), and no Block*Group interaction, \( F(4.21, 105.44) = 1.25, p = .27 \). These results provide no evidence of a change in subjective fluency in the Exp96 group, contrary to the predictions of hypothesis 4.
Table 4.15. Descriptive statistics for the subjective fluency ratings during the training phase in the 96-exposure condition by group and test item type.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>2.36</td>
<td>3.92</td>
<td>1.00</td>
<td>0.26</td>
<td>1.78 – 2.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.48</td>
<td>2.33</td>
<td>0.67</td>
<td>0.17</td>
<td>2.09 – 2.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.35</td>
<td>1.92</td>
<td>0.63</td>
<td>0.16</td>
<td>1.99 – 2.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.24</td>
<td>1.67</td>
<td>0.55</td>
<td>0.14</td>
<td>1.92 – 2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.27</td>
<td>2.50</td>
<td>0.78</td>
<td>0.21</td>
<td>1.82 – 2.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.57</td>
<td>2.83</td>
<td>0.88</td>
<td>0.23</td>
<td>2.06 – 3.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2.67</td>
<td>2.67</td>
<td>0.74</td>
<td>0.20</td>
<td>2.24 – 3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.64</td>
<td>2.50</td>
<td>0.76</td>
<td>0.20</td>
<td>2.20 – 3.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.23</td>
<td>3.25</td>
<td>0.74</td>
<td>0.19</td>
<td>-0.66 – 0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>2.37</td>
<td>2.50</td>
<td>0.72</td>
<td>0.20</td>
<td>1.93 – 2.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.34</td>
<td>2.25</td>
<td>0.85</td>
<td>0.23</td>
<td>1.83 – 2.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.58</td>
<td>2.08</td>
<td>0.74</td>
<td>0.20</td>
<td>2.13 – 3.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.67</td>
<td>2.08</td>
<td>0.73</td>
<td>0.20</td>
<td>2.23 – 3.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.24</td>
<td>2.92</td>
<td>0.75</td>
<td>0.21</td>
<td>1.78 – 2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.34</td>
<td>2.75</td>
<td>0.72</td>
<td>0.20</td>
<td>2.24 – 3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2.46</td>
<td>2.50</td>
<td>0.73</td>
<td>0.20</td>
<td>2.02 – 2.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.50</td>
<td>2.06</td>
<td>0.69</td>
<td>0.19</td>
<td>2.08 – 2.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.12</td>
<td>2.14</td>
<td>0.55</td>
<td>0.15</td>
<td>-0.46 – 0.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 4 predicted that Exp48 and Exp96 participants would acquire lower-order explicit knowledge as measured by subjective fluency ratings. It was also predicted that, due to the subtle nature of changes in fluency (Perruchet, 2008), participants in the 18-exposure condition would not experience changes in subjective fluency. The results reported in section 4.3.1.3 generally confirm this latter assertion. Overall, Exp18 participants were not significantly different from the Con18 participants in their subjective fluency ratings, although the Exp18 participants did show a small, but significant increase in perceived difficulty of reading the exposure phase sentences, contrary to Hypothesis 4. Moreover, participants in the Exp48 and Exp96 groups showed no evidence of changes in subjective fluency towards easier reading. In fact, the data show trends
toward the opposite: with increased exposure, participants tended to rate sentences as subjectively more difficult to read, although this trend was not statistically significant. Overall, the results do not support hypothesis 4. There is no evidence that participants acquired lower-order explicit knowledge as measured by changes toward increased ease of reading in subjective fluency ratings.

Research question #2 asked whether learners would acquire conscious knowledge of L2 syntax in the form of declarative memories, microrules, and/or changes in subjective fluency. In general, the results showed that learners did develop conscious knowledge of L2 syntax in the form of declarative memories and microrules, but not in the form of changes to subjective fluency. In terms of the specific hypotheses made at the outset of this chapter, the following results were obtained: Hypothesis 2 was fully supported. Hypothesis 3 was mostly supported. The Exp48 group did not have recognition memory for Early sentences, but all other Experimental groups did have recognition memory for all other types of test sentence. Hypothesis 4 was rejected.

4.3.3. Research question #3

In response to research question #3, hypothesis 5 predicted that declarative memory for L2 syntax would be passively memorized throughout the exposure phase. This section synthesizes some of the results reported above for the purpose of clearly evaluating hypothesis 5. If participants continued to form declarative memories during the exposure phase, then they should have recognition memory for recently-presented sentences. This hypothesis would be supported if Experimental participants could discriminate Old vs. New items in the 18-exposure phase, and Late items in the 48- and 96-exposure phases. Recognition memory difference score analyses
revealed that Exp18 participants were able to discriminate Old vs. New sentences better than the Con18 group, \( t(27) = 3.31, p = .003, d = 1.22 \). Likewise, Exp48 participants were able to discriminate Late Old vs. New sentences better than the Con48, \( U = 20.00, z = -3.45, p < .001, r = .66 \), and the Exp96 participants were able to discriminate Late Old vs. New sentences better than the Con96, \( t(25) = 2.13, p = .04, d = .85 \). To examine whether participants’ ability to discriminate Old vs. New for recently presented sentences changed with increased exposure, an ANOVA was performed on difference scores for the Exp18 group (\( M = 1.09, SD = 1.01 \)) and difference scores on Late items for the Exp48 (\( M = 1.78, SD = .82 \)) and Exp96 groups (\( M = 1.05, SD = .91 \)). The ANOVA revealed no significant between-group effect, \( F(2, 42) = 2.72, p = .08 \).

Since the overall ANOVA was not significant, post-hoc tests are not reported. In sum, the results show that declarative memories for L2 syntax were passively memorized during the course of the exposure phase, consistent with hypothesis 5.

4.3.4. Research question #4

Research question #4 asked whether declarative memory for L2 syntax from early in training would deteriorate over the course of the exposure phase. In response to research question #4, hypothesis 6 made two competing predictions: Hypothesis 6a predicted that if the ability to discriminate Old vs. New sentences was tied to the surface form of sentences from the exposure phase, then recognition memory for sentences presented in the first 18 trials of the exposure phase (Early items) should fade with increased input, i.e., because of the amount of time that had elapsed. Hypothesis 6b predicted that if the ability to discriminate Old vs. New sentences was based on memory for syntax that was not strictly tied to the surface structure of exposure phase sentences, then recognition memory for Early sentences should not fade with increased input.
In order to address research question #4, recognition memory difference scores for Early items were analyzed in the 48- and 96-exposure condition. Difference score analyses revealed that the difference scores on Early items in the Exp48 group ($M = .85, SD = 1.09$) were larger than those in the Con48 group ($M = .25, SD = 1.67$), but this difference was not significant, $U = 64.50, z = -1.29, p = .20$. This result suggests that declarative memory for Early items had deteriorated with increased exposure, consistent with hypothesis 6a. On the other hand, difference score analyses revealed that the difference scores on Early items in the Exp96 group ($M = 1.54, SD = 1.60$) were significantly larger than those in the Con48 group ($M = 0.46, SD = 1.28$), $t(25) = 2.00, p = .05, d = .76$. This result suggests that declarative memory for Early items had not deteriorated with increased exposure in the Exp96 group, consistent with hypothesis 6b. Making matters more complex, the difference scores on Early items in the Exp48 and Exp96 groups were not significantly different, $t(26) = 1.27, p = .214$.

In sum, the results do not provide clear support for hypothesis 6a or 6b. There are good grounds for rejecting hypothesis 6a, since the Exp96 group had recognition memory for Early items and the Exp48 group did not. If anything, hypothesis 6a predicts that the Exp48 group should have had superior recognition memory for Early items than the Exp96 group, since memory for L2 syntax was expected to deteriorate. On the other hand, hypothesis 6b cannot clearly be supported, because it predicts that participants should have had recognition memory for Early sentences in both the Exp48 and Exp96 groups. What do these results mean? Three plausible explanations are immediately available. The first, less theoretically interesting, possibility is that small sample sizes and large amounts of variation in performance on the recognition memory task have obscured true recognition memory effects in the Exp48 group. The second, more theoretically interesting, possibility is that both hypotheses are partially correct.
and there is some change from lexically-specific processing to non-lexically-specific processing that occurred during the course of the exposure phase. There are good theoretical and empirical reasons that this may have been so (e.g., N. Ellis, 2005; Goldberg, 1995, 2006; Tomasello, 2003; Ullman, 2004, 2005). For example, if early learning is lexically-driven, then one might expect more interference between Old and New items, because they are all equally old lexically. Increased training might lead to better abstraction and/or change learners’ heuristics making Old and New test items no longer equally old. This explanation is speculative, however, and more research is needed to better interpret the pattern of results on Early items in terms of theories of SLA.

A third option is that the presentation of Late items before Early items in the recognition memory test phase confounded the results. Participants always classified Late items before Early items in the recognition memory test. Now, given that pilot work showed possible interference effects during longer test phases, it is possible that the Exp48 group suffered from interference effects. This interpretation is particularly compelling if we consider that after just 48 total exposures, the Exp48 group may not have had robust memory representations for the structure of the stimuli. If this is so, it makes sense that they would perform well on Late items, but not Early items due to a combination of forgetting and test phase interference. Such interference would be less likely in the Exp96 group, whose memory representations for sentence structure may have been more resistant to interference. As such, it is not fully clear whether the decay of recognition memory for Early items is due to methodological flaws or to some internal processes that have yet to be adequately captured by theory. Future research is necessary to address the possibility of this confound.
Chapter V: Discussion

5.1. Overview of research questions and results

At the outset of this dissertation, it was noted that SLA researchers have a long-standing interest in the effectiveness of providing learners with conscious knowledge through explicit instruction (e.g., Sanz & Morgan-Short, 2004); however, there has been little investigation of how conscious knowledge develops under incidental learning conditions, without pedagogical intervention. Therefore, the aim of the present dissertation to investigate the development of conscious knowledge of L2 syntax under incidental learning conditions, focusing on: (a) the development of conscious L2 knowledge over time and (b) the representation of that conscious L2 knowledge. In what follows, I briefly summarize and contextualize the findings from Experiments 1 and 2. This is followed by a discussion of the wider theoretical and pedagogical implications of the dissertation. This chapter concludes with a consideration of the limitations of the studies reported here and avenues for future research.

5.1.1. Experiment 1

Experiment 1 asked three interrelated questions about the representation of L2 syntax acquired under incidental learning conditions. With respect to the first research question, Can statistical learning of L2 (i.e., semiartificial) syntax be obtained under incidental learning conditions? it was hypothesized that there would be evidence of statistical learning of the three syntactic structures. The results demonstrated no evidence of learning via changes in participants sentence-reading times; however, there was some evidence of a small learning effect of 2/3
target syntactic structures. In a grammaticality judgment task, Experimental participants were better at classifying structures A (TSPOV) and B (TSOPV) than Control participants, were worse than Controls on structure C (TVSPO), and were not significantly different from Controls on ungrammatical structures D (TSPVO), E (TSVPO), and F (TVSOP).

The second research question asked whether there was any behavioral evidence of statistical computation by reaction time measures. Hypothesis 2 predicted evidence of statistical computations in the form of reading time speedup for high-probability sequences relative to low-probability sequences. There was no significant evidence of such a speedup for high-probability syntactic category bigram sequences or for high-probability whole sentences. Therefore, the behavioral measures yielded no significant evidence that Experimental participants were performing statistical computations.

As for the third research question, Is the resulting knowledge consistent with a statistical computation view of statistical learning (as modeled by the SRN) or with a chunk formation view of statistical learning (as modeled by PARSER)? it was predicted that PARSER would better account for the human patterns of performance across the different structures in the grammaticality judgment task. This prediction was made in favor of PARSER because Experimental participants in the behavioral experiment demonstrated no evidence of statistical computations along the lines expected by the SRN. This hypothesis was broadly supported. PARSER was better able to account for the Experimental group’s pattern of performance across the different grammatical and ungrammatical structures at test. However, both the SRN and PARSER were actually better able to classify test structures than human participants, and neither model was able to fully reproduce the full range of human results. Therefore, the goodness of fit
of PARSER was relative to the SRN, indicating that other computational models (or other ways of training PARSER) will be necessary to account for the human data.

5.1.2. Experiment 2

Building on the findings of Experiment 1, the second experiment asked four interrelated questions about the development of conscious knowledge of L2 syntax over time. Research question #1 asked *At what point during incidental learning do learners initially develop conscious knowledge of L2 syntax?* Hypothesis 1 predicted that Experimental participants would acquire conscious knowledge of L2 syntax by recognition memory and verbal report measures, but not subjective fluency rating measures, after just 18 exposures (six exposures to each structure). Analyses of retrospective verbal reports showed that 14/15 participants became aware of some syntactic regularities after just 18 exposures, most commonly indicating conscious knowledge of the verb final regularity in structures A (TSPOV) and B (TSOPV). Likewise, recognition memory judgments indicate the presence of conscious knowledge of L2 syntax even when influences due to higher-order explicit knowledge (from the verbal reports) and implicit knowledge (in the form of priming) were controlled for, indicating the presence of genuine declarative memory for exemplars. As predicted, there was no evidence of conscious knowledge via changes in subjective fluency ratings. Overall, the results supported hypothesis 1.

Research question #2 asked *Do learners acquire conscious knowledge of L2 syntax in the form of declarative memories, microrules, and/or changes in subjective fluency?* Hypothesis 2 predicted that Experimental participants would acquire conscious knowledge in the form of higher-order explicit knowledge (e.g., microrules), regardless of the duration of their exposure phases. Hypothesis 3 predicted that Experimental participants would acquire conscious
knowledge in the form of declarative memories. Hypothesis 4 predicted that Experimental participants would acquire conscious knowledge in the form of changes in their subjective fluency ratings over the course of the exposure phase. In general, the results showed that learners did develop conscious knowledge of L2 syntax in the form of declarative memories and microrules, regardless of the duration of the exposure phase; however, there was no evidence of conscious knowledge in the form of changes to subjective fluency. Hypothesis 2 was fully supported. Hypothesis 3 was mostly supported. The Exp48 group did not have recognition memory for Early sentences, but all other Experimental groups did have recognition memory for all other types of test sentence. Hypothesis 4 was rejected.

Research question #3 asked whether Experimental participants would continuously form declarative memories for L2 syntax during the exposure phase. Hypothesis 5 predicted that declarative memory for L2 syntax would be continuously formed during the exposure phase. Results showed that regardless of the amount of input and training duration, Experimental participants were able to discriminate Old vs. New sentences on the recognition memory test (there were no significant effects of higher-order explicit knowledge or implicit priming). Thus, hypothesis 5 was supported.

Research question #4 asked whether declarative memory for L2 syntax from early in training would deteriorate over the course of the exposure phase. In response to research question #4, hypothesis 6 made two competing predictions: Hypothesis 6a predicted that if the ability to discriminate Old vs. New sentences was tied to the surface form of sentences from the exposure phase, then recognition memory for sentences presented in the first 18 trials of the exposure phase (Early items) should fade with increased input, i.e., because of the amount of time that had elapsed. Hypothesis 6b predicted that if the ability to discriminate Old vs. New
sentences was based on memory for syntax that was not strictly tied to the surface structure of exposure phase sentences, then recognition memory for Early sentences should not fade with increased input. The results of the recognition memory test did not provide clear support for hypothesis 6a or 6b. Hypothesis 6a should be rejected, since the Exp96 group had recognition memory for Early items and the Exp48 group did not, and, if anything, hypothesis 6a predicted that the Exp48 group should have had superior recognition memory for Early items than the Exp96 group, since memory for L2 syntax was expected to deteriorate. On the other hand, hypothesis 6b cannot clearly be supported, because it predicted that participants should have had recognition memory for Early sentences in both the Exp48 and Exp96 groups.

In the remaining sections, a discussion of the theoretical and pedagogical implications of the present studies is provided, and this is followed by an overview of the limitations of this dissertation and goals for future research.

5.2. Implications

In this section, implications of the present study are reported in the same order as theories were presented in the literature review: memory/usage-based approaches, incidental/implicit learning research, instructed SLA, and awareness research in SLA. Needless to say, these interpretations should be treated with caution, especially in light of the limitations of Experiments 1 and 2 discussed in section 5.3.

5.2.1. Implications for memory-based and usage-based accounts of SLA

Recall that both the Declarative/Procedural model (Ullman, 2004, 2005) and the memory model of Paradis (2004, 2009) predict that the earliest phases of syntactic development in adults will
rely on declarative memory. The results of Experiment 1 and 2 are consistent with this view. In Experiment 1, human data was better simulated by PARSER, a computational model that forms chunks. The basic principles of attention, awareness, and associative memory guiding PARSER mimic those implicated in various aspects of declarative memory (e.g., Ullman, 2004), suggesting the human data may have been driven by comparable declarative memory mechanisms. In Experiment 2, most Experimental participants formed higher-order explicit knowledge and had recognition memory for exposure phase sentences, both of which are posited to rely on declarative memory. It appears as though declarative memory for syntactic information is a consequence of simply reading sentences for meaning. Importantly, it remains to be seen how the mechanisms and neural substrates of declarative memory supporting recognition memory and higher-order explicit knowledge differentially contribute to these types of conscious knowledge, since they were shown to dissociate in Experiment 2.

The results of Experiments 1 and 2 were also consistent with usage-based accounts of SLA (e.g., N. Ellis, 2002, 2005, 2006, 2008; Ellis & Cadierno, 2009; Ellis & Larsen-Freeman, 2009; MacWhinney, 2008; Robinson, 1996, 1997; Robinson & N. Ellis, 2008; Roehr, 2008) and both usage-based and lexicalist accounts of syntax (e.g., Culicover & Jackendoff, 2006; Goldberg, 2006; Hagoort, 2005; Langacker, 1987; Vosse & Kempen, 2000; Tomasello, 2003). Broadly speaking, these approaches tend to place syntactic representations in lexical (declarative) memory, often stored as prefabricated, abstract chunks (e.g., constructions, N. Ellis, 2008; Goldberg, 2006; Robinson, 1996, 1997). These approaches do not rule out the possibility of combinatorial rules but downplay any strong roles for symbolic rules as they are traditionally conceived of in generative approaches to syntax (i.e., they reject the notion of unpronounced copy layers of syntax). Instead, most of these approaches only invoke a single combinatorial
process which combines constructions with other constructions, this is known as Unification in Culicover and Jackendoff’s (2006) and Hagoort’s (2005) frameworks. Interestingly, Unification is comparable to the Merge function in Minimalist syntax (e.g., Chomsky, 1995), suggesting that usage-based, lexicalist, and generative accounts are beginning to converge on the notion of syntax in the lexicon.

Now inasmuch as Experiments 1 and 2 both reported that syntax seemed to be processed by declarative lexical memory processes, then the present results are consistent with accounts of SLA and syntax that place the burden of syntactic development—at least in its early phases—on the lexicon and the memory system underlying it. Such a view is grounded in well-established processes associated with the neural substrates of declarative memory. For example, applied to learning complex structures, the hippocampal system automatically processes the surface structure, binding primitive units into more complex chunks, which are subsequently operated on by implicit learning mechanisms (N. Ellis, 2005, p. 317; Perruchet & Vinter, 2002; Schmidt, 1990, 2001; Ullman, 2004).

5.2.2. Implications for incidental learning research

The kinds of knowledge acquired in both experiments are consistent with previous findings in the incidental learning and implicit learning research traditions. First, there was no evidence for incidental learning of implicit, abstract rule knowledge (e.g., there was no categorical performance, indicating the application of a rule), consistent with previous work (e.g., Johnstone & Shanks, 2001; Perruchet & Pacteau, 1990; Rebuschat, 2008; Redington & Chater, 1996; Shanks & Johnstone, 1999; Williams & Kuribara, 2008). However, there was evidence of rule learning in the form of higher-order explicit knowledge (i.e., verbalizable metalinguistic
microrules), which is consistent with some previous work (e.g., Dulany et al., 1984, 1985; Mathews et al., 1989; Rebuschat, 2008).

There was also no behavioral evidence that learners performed unconscious statistical computations, at least not the ones embedded in the syntactic structures. In Experiment 1, reaction time measures showed that participants’ reading times did not significantly change in accordance with the statistical structure of the input (at least, not the transitional probabilities embedded in the stimuli). Participants also were not more accurate at classifying high-probability structures (transitional probabilities across structures: $M_A = .58, M_B = .42, M_C = .50$) on the grammaticality judgment task, contrary to what would be predicted by a statistical computation approach. Likewise, Experiment 2 showed no evidence that participants’ recognition memory corresponded with high-probability structures. That is, the formation of declarative memories did not appear to be driven by the transitional probabilities embedded in the input. These findings are consistent with previous research showing that statistical computations are unnecessary in accounting for human sensitivity to statistical structure (e.g., Giroux & Rey, 2009; Perruchet & Peereman, 2004). For example, Giroux and Rey (2009) showed that Aslin et al.’s (1998) and Saffran et al.’s (1996) demonstrations of statistical learning—which are widely regarded as evidence for statistical computations—are actually better captured by chunk formation processes, which are ubiquitous in the incidental and implicit learning literatures. Such chunk formation processes were also evident in the present experiments. When human performance in Experiment 1 was compared with a SRN (which learns via statistical computation) and PARSER (which learns via chunk formation), the patterns of performance on the different test structures were

40 Of course, prior knowledge may bias participants to certain aspects of the stimuli, changing the way their attention interacts with the statistical regularities (Zhao, Al-Aidroos, & Turk-Browne, in press).
better simulated by PARSER. This result was interpreted as evidence that similar chunk formation processes were active in the human participants.

In sum, the results of the present studies are consistent with the previous studies (outlined in Chapter 2) in SLA and cognitive psychology that employed incidental learning conditions. Evidence was obtained of fragmentary verbalizable metaknowledge (cf. Dulany et al., 1984, 1985; Hama & Leow, 2010; Rebuschat, 2008; Williams, 2005) and chunk formation (cf. Dienes et al., 1991; Giroux & Rey, 2009; Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990’ Perreuchet & Peereman, 2004; Perruchet & Vinter, 1998; Robinson, 2005, 2010) as was evidence of the incidental formation of declarative memories (e.g., Morgan-Short et al., 2012; Perruchet et al., 1997; Shanks & Johnstone, 1999). However, these types of knowledge were tapped differentially by different tasks, and more work is needed to clarify the extent to which higher-order explicit knowledge, chunks, and recognition memory derive from a single underlying knowledge base or from multiple different sources of knowledge. Moreover, more work is also needed to assess the extent to which other, perhaps implicit, mechanisms were involved in the learning processes reported here.

5.2.3. Implications for instructed SLA research

The key finding of this dissertation was that when learners simply read sentences with non-native syntax for meaning, they automatically began forming conscious knowledge about those sentences, represented as declarative memories and higher-order explicit knowledge. What is more is that learners appeared to continuously form such conscious knowledge throughout the exposure phase, regardless of its duration. These results fit what N. Ellis (2005, p. 320) calls the continuing development of “concrete seeds” which ultimately give way to “abstract trees.”
Encountering a novel or unexpected structure disrupts automatic implicit processing and appears to recruit resources necessary for the formation of conscious knowledge (N. Ellis, 1996, 2005; Haider & Frensch, 2005). These structures are accumulated over time (Robinson, 1996, 1997), while implicit mechanisms work to abstract across their patterns.

How does this finding address practical concerns? Well, for one it suggests that L2 learners are likely to develop their own conscious knowledge, even in the absence of instruction to do so. On the other hand, the results (although short-term in scope) showed that learners’ conscious knowledge did not change over the course of the exposure phase. More exposure did not lead learners to develop more accurate conscious knowledge (by recognition memory or verbal report measures)—or even more conscious knowledge generally (i.e., participants in the Exp96 condition did not report a wider variety of higher-order explicit knowledge than participants in the Exp18 and Exp48 groups). Consequently, learners left to their own devices did not appear to increasingly develop their conscious knowledge. This raises concerns for approaches that leave learners to discover their own rules, even simple ones (Doughty & Williams, 1998, p. 225). However, this conclusion is limited by at least two factors. First, the primary measures of learning focused on conscious knowledge. Different implicit measures may have revealed some evidence of increasingly robust development over time. Second, the structures used were not very diverse and were relatively simple, so participants may not have developed increasingly more complex knowledge because the materials were not sufficiently cognitively demanding (cf. Robinson & Gilabert, 2007). At any rate, teachers and instructed SLA researchers need to give due consideration to the ways in which learners develop their own conscious knowledge. It may be that learners’ apparently spontaneous creation of conscious knowledge in the face of novel L2 structures (N. Ellis, 2005, p. 308) would circumvent the need
to explicitly teach such structures in the classroom. On the other hand, learners may be prone to developing inaccurate conscious knowledge (e.g., over-generalizing the verb-final pattern in the present experiments), or may only develop conscious knowledge for certain types of forms, in which case targeted pedagogical interventions may be necessary. Given the brief scope of the present learning experiments, it is tempting, perhaps, to let learners continue to discover structure from more input in a more naturalistic second language learning situation; however, in the foreign language classroom, it may be more effective to use pedagogical interventions, since, as was demonstrated in Experiment 2, learners do rapidly develop conscious knowledge.

Another issue concerns learner attention during the experiments themselves. A qualitative look at the retrospective verbal report interviews reveals that participants’ attention was biased towards the edges of sentences (e.g., the first and last words or phrases). In fact, the bulk of participants’ verbalizable knowledge was related to the edges of the sentences. This finding should hardly come as a surprise. Attentional biases for edges are probably innate, and are found in human and non-human primates (e.g., Endress, Carden, Versace, & Hauser, 2009). This type of bias has been implicated in language acquisition: Slobin (1973) argues that attention is biased towards the edges of words, typically due to their informativity. Thus, it appears that adult L2 learners may be biased toward processing at the edges of their input. What does this mean for instructed SLA research? It is difficult to say. Obviously not all target L2 forms can be manipulated such that they conveniently occur at the end of exemplar sentences. However, a more appealing alternative is to make target forms that are not at the edges of sentences more salient: make them more likely to attract attention. This is not a new thought; after all, focusing learner attention is the goal of explicit instruction, recasts and other forms of corrective feedback, and input enhancement. Unfortunately, the present work does not offer a concrete solution for
learners’ attentional biases; however, the present work does reiterate that such biases can be problematic, and bolsters the need for instructional techniques designed to counteract or benefit from those biases.

5.3. Methodological innovations

Between Experiments 1 and 2, this dissertation included three methodological innovations in design that need to be considered in terms of their validity and future use in SLA research. This section briefly reviews these methods.

5.3.1. Triangulating awareness

Following calls for more triangulation in SLA research on awareness (e.g., Leow, 2000; Robinson et al., 2012), Experiment 2 employed three measures of conscious knowledge: recognition memory, retrospective verbal reports, and subjective fluency ratings. One of the many aims of using multiple measures of awareness is to raise the overall construct validity of the study. The construct of awareness and conscious knowledge is difficult to measure, and the lack of awareness is much more difficult to measure.

The purpose of the three measures employed here was to assess descriptively different types of conscious knowledge at different points of learning (but not during learning). For example, Experiment 2 demonstrated a dissociation between the higher-order explicit knowledge that participants verbalized in the retrospective verbal report interviews and their recognition memory performance. Once the internal validity of the recognition memory measures has been raised (e.g., by controlling for other factors, such as higher-order explicit knowledge and implicit priming), recognition memory serves as a useful measure for conscious knowledge, especially in
paradigms where the contents of consciousness may be difficult to describe verbally (e.g., SRT tasks, Perruchet et al., 1997). Indeed, Experiment 2 showed that participants’ verbal reports underdetermined their performance on the recognition memory task, suggesting that recognition memory judgments were made on the basis of knowledge that was either (a) functionally distinct from that which supported their verbal reports, or (b) difficult to verbalize.

The present experiments do not allow conclusions regarding which of these solutions is correct (a, b, both, or neither); however, this does provide a promising avenue for future research in SLA. For example, it may be that performance on recognition memory tests and retrospective verbal reports is subserved by the same underlying memories for the training material. On this view, it is the difficult-to-express qualities of the memories that give rise to the dissociations on these different measures of awareness. Such a finding would be broadly consistent with single-system views of declarative memory function (e.g., Shanks & Berry, 2012). On the other hand, the present dissociation between recognition memory and higher-order explicit knowledge may represent truly different knowledge bases, with the former relying on hippocampal and temporal lobe structures and the latter involving frontal lobe structures (N. Ellis, 2005, p. 317). Either way, understanding the underlying memory mechanisms that support descriptively different types of conscious knowledge will be a crucial step in ultimately understanding the relationship between L2 development, the development of conscious knowledge, and the optimal role of explicit pedagogical intervention.

5.3.2. Trained control groups

This study continued the use of trained control groups from previous research (e.g., Hamrick, 2012, 2013; Rebuschat et al., forthcoming). The purpose of trained control groups is to make the
degree of noise (i.e., non-learning related variations in performance) as maximally similar in Experimental and Control groups as possible. This procedure has been argued to maximize the likelihood of finding genuine learning effects and avoiding false-positive evidence of learning (e.g., Hamrick & Sachs, in preparation; Perruchet & Reber, 2003). In short, trained control groups should, in theory, prevent fluke performance factors. In Experiments 1 and 2, the use of trained control groups revealed genuine learning effects in the Experimental groups; however, these learning effects were not as robust as they would have appeared when compared to chance as a baseline (as has been done in previous research, e.g., Williams, 2005) or when compared to untrained controls (e.g., Rebuschat, 2008). For example, the Experimental group in Experiment 1 performed significantly above chance overall, $t(12) = 2.25, p = .04$.

All of these considerations depend on the function of the control group. Using chance and untrained controls as a baseline run the risk of non-learning-based influences on performance in experimental groups going undetected. As such, appropriately trained controls should be designed to have the same non-learning-based influences as the experimental groups as much as possible. However, it is important to keep in mind that training controls on all the same tasks, but with randomized or pseudorandomized stimuli, may create a bias towards randomness (or perceived randomness). Thus, rather than controlling for non-learning-based noise, one has simply replaced one bias (i.e., towards the experimental stimuli) with another (i.e., towards randomized stimuli). This may result in a trained control group that is not neutral, but biased differently, leading to more uncertainty in the ultimate interpretation of the data. Keeping all this in mind, it is hoped that the studies reported here will encourage the use of appropriately trained control groups (in addition, perhaps, to an untrained control group) in future research in SLA in
order to more clearly elucidate the nature of adult L2 learning capabilities and avoid misleading results.

### 5.3.3. Comparing multiple computational models of learning

Previous research has typically relied on a single model or class of models to replicate human behavior. For example, Ellis and Schmidt (1997) and Williams (2010; Williams & Kuribara, 2008) have used connectionist networks to replicate human patterns of behavior on naturalistic and laboratory L2 learning phenomena. However, these studies did not compare the results of multiple models that make orthogonal or competing predictions. Instead, the simulations acted as proofs of concept. Connectionist networks can, in principle, simulate an incredible range of human behavior when coded appropriately. However, a far more compelling demonstration of a fundamental similarity between a computational model and human cognition is derived from training said model across a large parameter space and comparing it with other competing models. These two innovations (insofar as we are focused on SLA research) were performed in Experiment 1.

First, it was crucial that the SRN and PARSER were both trained across a wide range of parameters. This ensured that any goodness-of-fit between either model and human behavior was due to the general architecture and principles of the model, and not due to some idiosyncratic method of performing the simulations or coding the stimuli (e.g., Boucher & Dienes, 2003). Second, multiple models were compared. Both the SRN and PARSER are considered classes of statistical learning models; however, they operate on very different principles and have been shown to develop preferences for different types of statistical information (e.g., SRN becomes sensitive to forward transitional probabilities; PARSER becomes sensitive to mutual correlation...
statistics). If we ultimately want the best model that can account for the widest range of human data, then it will be crucial in future work to continue the trend of comparing multiple models. Better yet, it would be ideal to conduct simulations across multiple models that make a wide variety of assumptions and are not just statistical learning devices (e.g., ACT-R, Anderson, 1996).

5.4. Limitations and future research

The experiments reported in this dissertation contained a number of limitations, each of which must be addressed if future research is to further clarify the development of conscious knowledge during incidental L2 learning.

First and foremost, the experiments are limited by the use of semiartificial language stimuli in an artificial setting (self-paced reading on a computer). As a consequence, participants in Experiments 1 and 2 may have been activating complexes of syntactic knowledge during lexical retrieval which then had to be suppressed in order to process the L2 syntax. This problem may have been one contributing factor to the wide standard deviations reported in the present studies (especially when combined with the small sample sizes). It is easy to imagine that individual differences (or even learner-internal variations) in inhibiting or suppressing irrelevant L1 grammatical knowledge may have led to the large variations in performance. This problem may be overcome by using a natural language for input in a more naturalistic setting, but such a procedure may not ultimately have avoided some of the present limitations. For example, the artificial introduction of statistical patterns to a natural language would have made it (no surprises) more artificial. The statistics in natural language are not all 67% and 33%.

Consequently, researchers should take into consideration their ultimate aims in deciding on their
choice of stimuli. In the present experiments, it was important for practical reasons to use a semiartificial language in order to obviate the need for vocabulary pretraining. Moreover, since the semiartificial language was already functionally artificial, adding specific statistical patterns did not introduce a drastic change in the overall “weirdness” of the stimuli. To investigate learners’ sensitivities to statistics, using a miniature artificial language (e.g., Brocanto 2, Morgan-Short et al., 2012) may be useful for future research, provided face validity is not lost and learners are motivated enough to learn a language that has no value outside the laboratory.

Another limitation concerns the fact that the syntactic structures in the semiartificial language were devoid of meaning. Although it remains an open theoretical question whether syntax itself has meaning (see Cullicover & Jackendoff, 2006; Goldberg, 2006), there is general consensus that syntactic differences lead to at least some changes in the meaning of sentences (with the amount varying by language and syntactic structure). However, no such meaning was available in the syntax of the semiartificial language, since the same compositional meaning of each core sentence was rotated around all three semiartificial syntax structures. In part, this was a strategic choice designed to keep the stimuli more controllable: adding meanings would have substantially complicated the stimulus materials. However, future research would do well to incorporate some meaningful differences between syntactic structures to investigate what is learned when syntax encodes meaning. For example, Casenhiser and Goldberg (2005) report learning of a meaningful word order pattern after just 16 exposures to semiartificial sentences encoding the pattern (e.g. The kind the spot mooped).

Although controlling the details of the stimuli was complex, the actual syntactic structures used were quite simple. The simplicity of the stimuli may have aided the development of conscious knowledge more than what might be found with more complex stimuli. That is, the
development of conscious knowledge might have been tacitly favored by the presence of simpler structures. The possibility that conscious knowledge formation is easier when stimuli are simple has been put forth elsewhere (e.g., DeKeyser, 1995; Reber, 1989, 1993). Future research should be carried out with more complex structures to determine the extent to which learners develop conscious knowledge (e.g., perhaps using more complex semiartificial language stimuli, such as those in Rebuschat, 2008).

Another limitation of the stimuli concerns the recognition memory test stimuli in Experiment 2. Recall that there were three Old structures (A, B, C) and six New structures (D, E, F, X, Y, Z). Six New structures were used instead of just three to make sure that recognition memory performance was not due to idiosyncratic rejection of a given structure or two. This posed no problem in the 18-exposure condition, where half of the participants saw New structures D, E, and F, and the other half saw X, Y, and Z. However, these different New structures were confounded with Early and Late items in the 48- and 96-exposure phase. That is, structures D, E, and F always occurred as Early-New items, while structures X, Y, and Z always occurred as Late-New items. This is an important limitation because differences in performance on Early-New and Late-New structures may not have been due to just the effect of time, but also the effect of using different structures for Early-New and Late-New items. It will be important in future work to unconfound these factors by first using a single set of New structures that does not change regardless of whether it is for Early or Late items.

A final limitation of the stimuli concerns how they were treated in terms of chunks in both theory and the computational models. At various points in this dissertation it was stated that chunks are difficult theoretical constructs to pin down. Most agree on certain facts about chunks, e.g., they encode sequential information. However, there is less consensus on other issues,
especially the issue of the abstractness of chunks. Since much of the theoretical work on chunking has been done in the artificial grammar learning paradigm, it has commonly been implied that chunks are made up of surface structure (although, as noted in Chapter 2, this does not mean that they cannot be implicit, Knowlton & Squire, 1994, 1996). This makes intuitive sense, since the symbols used in artificial grammar learning tasks do not have underlying abstract categories or meanings associated with them. The same cannot be said for language. The symbols of language (words and morphemes) are associated with a range of abstract category information, especially information regarding syntactic categories. As such, chunks of language have been argued to encode syntactic information, regardless of the amount of lexical content in those chunks (e.g., constructions such as How about X? and Verb-Noun Phrase are both considered chunks of syntax stored in memory). Such arguments have come from a range of theoretical approaches (e.g., generative linguistics, Culicover & Jackendoff, 2006; cognitive linguistics, Goldberg, 1995, 2006; psycholinguistics, Hagoort, 2005; Perruchet & Poulin-Charronnat, forthcoming; head-driven phrase structure grammar, Pollard & Sag, 1994), and therefore should not be treated as an idiosyncratic hypothesis of a single theory. At times in the present dissertation, chunks were treated in abstract terms (i.e., in terms of their syntactic categories) and at other times in terms of surface structure (i.e., the actual words in the semiartificial language sentences). While both ways of talking about chunks can be theoretically and empirically motivated, it would still be preferable in future work to take a more consistent stand on the representation of chunk knowledge and elucidate more clearly when chunks should be considered as units of abstract structure and when they should be considered as units of surface structure.
In addition to inherent limitations of the stimuli used in this dissertation, there were also shortcomings in the methodology used. For example, in Experiment 2 the number of test items in the recognition memory phase was very small (12 items for the E48 and E96 conditions and 6 items for the E18 condition). Pilot work showed that interference effects set in rapidly after a few test items (presumably due to the very subtle differences between stimuli). Using 24 test items showed accurate performance on the first 12-16 items but substantially lower performance on the last half (16-24) of the recognition memory test. This effect persisted regardless of the order of presentation of test items, suggesting that it was interference, not the individual test items themselves that caused decreases in accuracy. Thus, in order to avoid substantial interference effects, the decision was made to have the smaller number of observations in the recognition memory task. It will be important in future work to find a way around this problem, but since interference effects are rampant in recognition memory tasks, it may be that an altogether different, orthogonal measure of declarative memory will have to be used.

The present study was also limited by the number of participants who ultimately ended up in each group. For example, many groups in the present dissertation had just 13 or 14 participants each. This sample size is robust enough to establish learning effects on some types of tasks with high internal reliability, but they may not be robust enough to establish learning effects for measures that are inherently noisy, like reading time measures. Unfortunately, a priori power analyses were not conducted. This is a serious limitation on the generalizability of the present results, and it will be absolutely essential to do so in future work in order to gauge the reliability of nonsignificant findings. In the present dissertation, participants’ reading times typically had quite large standard deviations, and this should hardly be surprising given the nature of the task and the stimuli. The use of semiartificial languages—and their deployment in a
reading task—probably involve the activation and then suppression of a wealth of syntactic knowledge. Because participants’ individual differences in ability to suppress their irrelevant L1 knowledge will likely vary widely, results are also likely to vary widely. That said, it is possible that the differences between Experimental and Control groups across both studies would have been significant with the addition of more participants. It will be important for future research to determine whether the lack of robust significant differences in reading times between groups is the result of a genuine lack of learning, individual differences in suppression of the L1, or an unfortunate symptom of a smaller sample size.

Another methodological flaw was in the use of the subjective fluency ratings. Although the use of these ratings was rooted in theory (fluency-based heuristics, Servan-Schreiber & Anderson, 1990), the measure itself was novel and had not been extensively piloted prior to this dissertation. Making matters worse, since all three conditions of Experiment 2 were conducted simultaneously, there was no chance to jettison these measures when it was found that they were not revealing. On the other hand, the on-line reading time performance of the Experimental groups in Experiment 2 (see Appendix) showed no significant change in reading times over the course of the exposure phase in Experiment 2. Taken together, the subjective fluency ratings may be capturing participants’ accurate introspections that they did not appear to gain robust fluency during the exposure phase. That said, the spirit of the subjective fluency ratings (as on-line measures of lower-order awareness) is still compelling, and future research is needed into a reliable on-line measure of levels of awareness that are too subtle to be detected by more coarse-grained methods like think-alouds. Indeed, it may be that subtle wording changes would reveal a different picture (i.e., asking the participants to indicate how easy or difficult it was to comprehend each sentence as opposed to read each sentence). Finally, the subjective fluency
ratings, unlike the plausibility judgments used in Experiment 1 (cf. Rebuschat, 2008) did not require participants to actually read the sentences in order to make a response. Thus, it is possible that participants did not read every single sentence during the exposure phase. It will be important to add something like comprehension questions to each trial to ensure that participants are reading the exposure phase sentences.

The present study also reported no evidence of the influence of implicit knowledge on performance; however, only one measure of implicit knowledge was used: priming. While participants did not appear to use priming as the basis of their performance on the recognition memory task, they may well have used other sources of implicit knowledge that simply were not measured. Participants may have formed implicit knowledge of the syntactic structures that was slowly accessed, due to the fact that learning was in its early stages. This is consistent with Cleeremans’ (2007) Radical Plasticity Thesis, which holds that implicit knowledge develops early in learning and can be used to influence behavior, but does not operate automatically. Future research is necessary to illuminate this picture. Perhaps the inclusion of subjective measures of awareness (e.g., Rebuschat, 2008, who used confidence ratings and source attributions to determine the conscious status of knowledge) would provide additional information about the possibility that learners used implicit knowledge in the recognition memory task.

As a final methodological flaw, it is worth noting that a better system could have been used for assessing the contribution of higher-order explicit knowledge to the recognition memory judgments. Because it was impossible to determine, a priori, how many participants would demonstrate awareness of the syntactic patterns on the retrospective verbal report measures (let alone determine which kinds of verbalizable knowledge they would produce), statistical analyses
comparing knowledge from the verbal reports with recognition memory test performance had to be conducted after more qualitative analyses of the interview data were conducted. Moreover, the most commonly reported higher-order explicit knowledge in the verbal reports was the verb final regularity. As a consequence, analyses were often limited to contrasts in performance on +VF and –VF old and new items. While these analyses did demonstrate recognition memory performance above and beyond what would be expected by just using a verb final microrule, it remains an open question whether similar conclusions could be drawn for the other types of higher-order explicit knowledge participants verbalized. In other words, participants who reported higher-order explicit knowledge that a sentence could begin with a temporal phrase and a subject, and end with a verb may not have used recognition memory at test, but may have only relied on his/her higher-order explicit knowledge. Such a possibility was not assessed for every different kind of knowledge participants expressed. Thus, Experiment 2 provided evidence of genuine recognition memory, not attributable to higher-order explicit knowledge; however, it only did so for participants reporting the verb final pattern and nothing else. It remains to be seen whether participants who report more or different higher-order explicit knowledge in the verbal reports also perform similarly on recognition memory tasks, or whether their more robust higher-order explicit knowledge differentially influences performance.

The generalizability of the present study is also limited by the brevity of the exposure phase. Granted, the focus was on the early phases of syntactic development, and, so, it was important to take measures of conscious knowledge after very short training durations. However, even the longest training period only consisted of the presentation of 96 exemplars (32 of each structure), and even for a study interested in the earliest phases of learning, the duration of training is still small. Likewise, participants were tested immediately in all conditions. No
delayed post-tests were employed, and, therefore, it is impossible to say to what extent the present results would persist over hours, days, or weeks. Therefore, the developmental claims of this study are limited to immediate outcomes of brief exposure periods. Additional research with a longer training period, perhaps using a longitudinal design, is needed in order to clarify how learners continue to develop conscious knowledge during L2 acquisition.

5.4.1. Future research

In addition to addressing limitations in the present dissertation, future research is also needed to examine questions raised here. For example, it remains to be seen to what extent learners automatically develop conscious knowledge for different L2 forms and form-meaning mappings. After all, it is possible that learners have a more difficult time forming declarative memories for certain L2 structures. Likewise, it may be the development of higher-order explicit knowledge is also dependent upon the features of the target forms. For example, the verb final pattern was salient in Experiments 1 and 2 for a variety of reasons, and, moreover, since it was reliable, the development of higher-order explicit knowledge should come as no surprise. On the other hand, features that are inherently difficult to grasp, like those governing the English article system, may not lead to the spontaneous development of higher-order explicit knowledge—not just because they are difficult to be aware of, but because they are very difficult to verbalize (although they may be memorized as complex chunks). Seeing how conscious knowledge of forms of varying complexity is incidentally acquired will be of the utmost importance in determining the extent to which learners automatically develop conscious knowledge. Since it has been argued that form complexity may how influence learners will benefit from additional explicit instruction (e.g., DeKeyser, 1995; Robinson, 1996), it will be necessary to use a battery
of criteria to measure form complexity and determine (a) what kinds of conscious knowledge
learners develop on their own and (b) how pedagogical interventions can best complement what
learners discover autonomously.

One major avenue of research to be explored in the future is to investigate the cause of
the verb final bias that participants so often reported. Here I briefly discuss two possible causes.
First, it may be that task instructions guided learners to preferentially process verbs, leading them
to form microrules on the basis of those verbs. In Experiment 1, this meant that participants were
waiting for the last word of each sentence (67% of the time this was the verb) in order to
determine the plausibility of the sentence. In Experiment 2, that the exposure phase task
instructions asked participants to pay attention to the meanings of the sentences in the same way
that they would when reading a book, article, or blog. According to verbal reports, this led some
participants to adopt a strategy of attending especially to the subject noun phrase and verb phrase
during training. This leaves open the question of why there was still a strong preference to
verbalize microrules in terms of verbs, rather than both noun phrases and verbs, since the
location of the subject noun phrases were also reliable (occurring 67% of the time in the second
position).

Learners may also have been biased by another strategy in which they formed conscious
knowledge on the basis of natural attentional biases in processing. Slobin (1973) noted that an
attentional bias in infants towards the edges of words would lead to informative processing, since
the edges of words (especially the ends of words) often carry important inflectional grammatical
information. Therefore, adult L2 learners may be carrying a similar bias into the present
experiments. On the other hand, an attentional bias for the edges of stimuli may not be language-
specific. Comparable biases have been found in human and non-human primate learning of non-
linguistic stimuli. For example, Endress et al. (2009) demonstrated that both humans and chimpanzees rapidly form knowledge on the basis of positional information at the edges of sequences (e.g., $A$ and $F$ in the sequence ABCDEF). In our example, this means the position of $E$ would be encoded relative to $F$. Thus, participants may not just learn that $E$ goes into the second-to-last position, but, rather, immediately prior to $F$, regardless of the context.

This sensitivity to the edges of stimuli exists for both visual and auditory stimuli. This suggests that the edge bias in the present dissertation was likely not due to the stimuli being presented in the written modality. This bias also goes a long way toward explaining the paradoxical comments made about the temporal phrase in the exposure phase. For example, 11 Experimental participants reported that they recalled that the temporal phrase came first in sentences, but 5 of these participants also reported ignoring it, since it carried little information that helped them complete the task (i.e., it did not carry important information). As such, participants appear to have attended to this information long enough to form a generalization and subsequently ignore it. It would be interesting for future research to investigate how these various attentional biases and task instructions interact during the course of learning. Learners undoubtedly bring a number of attentional biases to any task (some innate, some learned on the basis of prior linguistic experience) and it remains to be seen to what extent task-essential practice (Loschky & Bley-Vroman, 1993), and corrective feedback can lead learners to overcome their, sometimes unproductive, attentional biases (VanPatten, 2002) that they displayed in the present experiments. A simple extension of the present experiments by independently manipulating the information at the edges of sentences and the information required for completing the task (subject noun phrases and verb phrases) may go a long way toward addressing some of these questions.
It would also be interesting to investigate the extent to which individual differences influence development of conscious knowledge. For example, it is plausible that individual differences in working memory, aptitude, and motivation may influence the development of higher-order explicit knowledge (e.g., Reber, 1993). However, it is less clear how individual differences in any of these capacities would influence lower-order explicit knowledge as measured by, say, recognition memory tests.

Another avenue for future research would be the exploration of whether learners used analogical processes (consciously or unconsciously) to perform the recognition memory test. Rather than passively memorizing fragments or chunks of sentences, participants may have only memorized a few fragments and used them to analogically process subsequent sentences in the exposure phase and/or test phase. There is some evidence that analogy promotes generalization in language acquisition (e.g., Gentner & Namy, 2006), but it remains to be seen to what extent it was active here.

Finally, it would be interesting in future research to investigate the extent to which the present phenomena persist in a more naturalistic setting. For example, it would be interesting to see if EFL learners reading a textbook, listening to authentic speech, or interacting with a native-speaking interlocutor form the same sorts of declarative memories and metalinguistic knowledge as participants did in the present dissertation. To the extent that the kinds of conscious knowledge formation reported here are important for early development, it would be interesting to see how different pedagogical interventions and different types of task demands mediate and/or promote the formation of such conscious knowledge.

In sum, although the present dissertation offers several interesting findings about the development of conscious knowledge of L2 syntax during incidental learning, the studies
reported here still suffer from several limitations that constrain the generalizability and interpretability of the findings. However, there are various future research topics that stem from these limitations each of which will clarify the phenomena discussed here.

5.5. Conclusion

In two experiments, the present dissertation demonstrated that it is possible for adult learners to acquire conscious knowledge of L2 syntax under incidental learning conditions. Computational simulations suggested that learners’ knowledge was better captured by a chunk formation than statistical computation, suggesting that learners may have been forming abstract chunks from their input. Experiment 2 revealed that learners developed declarative memories for the sentences they read as an automatic consequence of processing those sentences for meaning. The representation of those declarative memories must have contained enough syntactic information for Experimental participants to distinguish sentences that were all equally old in terms of their lexical items and semantics, but which differed only in their syntax. Finally, participants were able to verbalize some knowledge, but this was primarily limited to information about what occurred at the beginnings and endings of sentences. Although the present studies leave many lingering questions, the findings of this dissertation suggest that adult learners spontaneously develop conscious knowledge (in the form of declarative memories and metalinguistic knowledge) simply by processing sentences for meaning and that they do so from very early on in learning, and continue to do so with more input, although it remains unclear whether this conscious knowledge fades with time. From a methodological viewpoint, the fact that learners spontaneously form different kinds of conscious knowledge makes it ever more important that
future research utilize multiple methods to triangulate the relationships between awareness, memory, and L2 learning. The present dissertation represents a step in that direction.
Appendix A

Experiment 1 Materials

Training Instructions and Stimuli

This experiment is about how we comprehend the meanings of sentences.

In this study, you will read scrambled English sentences one fragment at a time as QUICKLY and ACCURATELY as possible.

Your task is to:
(1) Press the SPACE BAR to see each new sentence fragment.
(2) Repeat the sentence out loud when the prompt appears.
(3) Judge whether the MEANING of each sentence is plausible.

Press the SPACE BAR to continue.

During the experiment, you will read scrambled sentences one fragment at a time.

For each sentence, you will first see a fixation cross in the middle of the screen:

+  

Once you see the fixation cross, you may press the SPACE BAR whenever you are ready to start the new sentence. Continue pressing the SPACE BAR to advance through each fragment of the sentence.

Be sure to read the sentences as QUICKLY AND ACCURATELY as possible. Your speed and accuracy are being measured.

Press SPACE BAR to continue.

When you finish each sentence, you will be prompted to repeat it word-for-word, exactly the way you read it. Do your best to say it exactly the way it was written.

When you have repeated the sentence, press the SPACE BAR to advance to the next item.

As soon as you have finished each sentence, you will see a box that asks "Plausible?"

When you see this box, you must indicate whether you think the sentence's meaning makes sense. In other words, would this scenario be likely to happen in the real world?

Press the GREEN KEY if you think the sentence was plausible. Press the RED KEY if you think the sentence was not plausible.

You will have 5 seconds to judge plausibility.
Press the SPACE BAR to continue.
So remember, your task is to:

(1) Hit the SPACE BAR to read the scrambled sentences as QUICKLY and ACCURATELY as you can.

(2) Repeat the sentences out loud when asked to repeat.

(3) Indicate whether the MEANINGS of the sentences are PLAUSIBLE.

Remember to keep your fingers ready by having them near the appropriate keys!

When you are ready to practice, press SPACE BAR.

Structure A (TSPOV)
Earlier today\David\ in the bedroom\the lightbulb\changed
Yesterday\Patrick\ in the laboratory\the rat\noticed
This morning\Steven\ from the desk\the pencil\got
Yesterday\Carol\ to the car\luggage\carried
Last month\Richard\ in England\the vacation\cancelled
Some time ago\Charlie\ at the garage sale\blankets\sold
Earlier today\Mary\ on her wrist\the bracelet\put
Last spring\Sarah\ in the shed\the birdhouse\made
Yesterday\Michael\ at the concert\the guitar\played
This morning\William\ from the book\ideas\explained
Last week\Jason\ on the internet\classes\taught
This past year\Kaitlyn\ at the meeting\the necklace\wore
Last week\Kevin\ at the club\the crowd\entertained
Last month\Laura\ from the balcony\the waves\watched
Some time ago\Alice\ from the wreckage\the survivor\pulled
Earlier today\Nicole\ from the magazine\stories\read
Last month\David\ in the tube\the toothpaste\shattered
This morning\Patrick\ through his telescope\Saturn\dropped
Some time ago\Laura\ in her garage\the car\saved
Some time ago\Susan\ for a sandwich\bread\operated
Earlier today\William\ at the river\the fish\mended
Last week\Charlie\ on his hand\his fingers\discovered
This morning\Caro\ in counseling\her future\recalled
Earlier today\Kaitlyn\ in her living room\the peaches\gambled
This morning\Michael\ at the mall\his hat\sailed
Some time ago\Richard\ at his workstation\the mousepad\filled
Yesterday\Jason\ in the oven\the pizza\sketched
Earlier today
Steven\inhis heart\crushed
Last month\Mary\out of the door\the garbage\included
Last week\Alice\through the meadow\her dog\unloaded
Yesterday\Sarah\at the theme park\the roller coaster\pushed
Last month\Nicole\on the corner\the tree\justified

Structure B (TSOPV)
Some time ago\David\the plans\for the weekend\confirmed
Last week\Richard\clothes\in his hamper\washed
Earlier today\Patrick\money\for a coffee\needed
Last month\Kaitlyn\tennis\for the college\coached
Yesterday\Charlie\the checkbook\at the bank\left
Last month\Steven\literature\in England\studied
Last week\Laura\family\on vacation\visited
Some time ago\Nicole\the results\from the report\disputed
Some time ago\Mary\an email\to her sister\typed
This morning\Sarah\ideas\in her diary\wrote
Earlier today\Carol\questions\from her boss\answered
Yesterday\Jenny\the wood\with the axe\split
This morning\Jason\the employee\for hard work\praised
Yesterday\Alice\the concepts\from the seminar\understood
Last week\Susan\wine\with dinner\drank
This morning\Tracy\the presents\for her brother\forgot
Yesterday\Richard\the comb\in the cabinet\wore
Earlier today\Jason\the cushion\on the sofa\demolished
Some time ago\Steven\the wind\from the storm\found
Some time ago\Kaitlyn\stars\at night\cancelled
Earlier today\Sarah\the surfboard\to the ocean\explained
Last month\Carol\the eggs\for breakfast\changed
This morning\Jenny\the cells\in her brain\noticed
Earlier today\Tracy\water\on the moon\got
Last week\David\mercy\for his enemy\carried
This morning\Charlie\the liquid\from the bottle\read
Last month\Patrick\the love\in his heart\observed
This morning\Alice\nails\into the wood\played
Last week\Mary\the heat\from the sun\bought
Yesterday\Laura\jam\in the jar\entertained
Last week\Nicole\shapes\out of paper\sold
Yesterday\Susan\the pictures\in their frames\taught

Structure C (TVSPO)
Some time ago\heaved\Patrick\onto the table\the backpack
This morning\scrutinized\Laura\for the building\the blueprints
Some time ago\shattered\Carol\in the dining room\the plate
Earlier today\dropped\Jenny\on her foot\the can
This morning\gambled\Richard\at the racetrack\his savings
Earlier today\sailed\Alice\in the Atlantic\her boat
Test Instructions and Stimuli

The word order of the previous sentences was not arbitrary, but determined by a complex rule-system instead. In other words, we used certain rules to determine what word goes where.

In the second part of the experiment, you will read 36 new sentences.
Importantly, all of these new sentences are plausible. Their meanings make sense. The difference is that half of these new sentences were made with the same word-order rules as in the first part of the experiment. The other half of these new sentences were made to break the word-order rules of the first part of the experiment.

Your task is to read these new sentences one at a time and decide whether they follow the SAME WORD ORDER RULES or whether they BREAK WORD ORDER RULES from the sentences in the first part of the experiment.

Press the SPACE BAR to continue.

In this test phase, you will read 36 new sentences.

18 of the sentences were created by the same rule-system as in the first part of the experiment, and so the word order will be the same as the sentences in the first part. These sentences are called GRAMMATICAL.

The other 18 sentences DO NOT conform to the rule-system, and so the word order will be different from the sentences in the first part. These sentences are called UNGRAMMATICAL.

Your task is to decide which sentences are grammatical and which are ungrammatical.

Press the SPACE BAR to continue.

Just as in the previous part of the experiment, you will hit the SPACE BAR to read through sentence fragments. Be sure to read as quickly and accurately as you can.

When you reach the end of each sentence, you will be asked if it was grammatical. In other words, is the WORD ORDER based on the same word order as the sentences in the first part of the experiment?

Answer YES by pressing the green button.
Answer NO by pressing the red button.

Press the SPACE BAR to continue.
So, your task is to do the following:

1.) Hit the SPACE BAR to read the sentences as QUICKLY and ACCURATELY as you can.

2.) Judge the word order of each sentence and indicate whether the sentence is GRAMMATICAL or UNGRAMMATICAL according to the word order rules from part one.

Press the SPACE BAR when you are ready to begin.

A
Not long ago\Calvin\in the fridge\the pear\kept
Recently Edward on the street his Jeep parked
Not long ago Brandon from the pantry the soup retrieved
Recently Abby in the kitchen chili ate
Not long ago Rosie in the church the singers observed
Recently Vickie on the beach shells collected

B
Not long ago Vickie the invitation to the party accepted
Recently Calvin the student for the job appointed
Not long ago Edward the thief from the store chased
Recently Brandon Mars in orbit researched
Not long ago Abby air from her lungs expelled
Recently Rosie the flowers in the vase watered

C
Not long ago included Rosie in the secret her friend
Recently saved Abby from dying the bird
Not long ago unloaded Calvin from the truck the piano
Recently typed Edward from his chair a letter
Not long ago visited Brandon over the holidays his relatives
Recently understood Abby for the lesson the goal

D
Not long ago Abby in the fridge kept the pear
Recently Rosie on the street parked her Jeep
Not long ago Vickie from the pantry retrieved the soup
Recently Calvin in the kitchen ate chili
Not long ago Edward in the church observed the singers
Recently Vickie on the beach collected shells

E
Not long ago Brandon accepted to the party the invitation
Recently Abby appointed for the job the student
Not long ago Rosie chased from the store the thief
Recently Vickie researched in orbit Mars
Not long ago Calvin expelled from his lungs air
Recently Edward watered in the vase the flowers

F
Not long ago understood Edward the goal for the lesson
Recently included Brandon his friend in the secret
Not long ago saved Abby the bird from dying
Recently unloaded Rosie the piano from the truck
Not long ago typed Vickie a letter from her chair
Recently visited Calvin his relatives over the holidays
Appendix B

Experiment 2 Materials

Exposure Phase Instructions

Welcome, and thanks for participating in this study!

This experiment is about how adults read sentences under "unusual" conditions.

In this study, you will be reading English sentences that have been scrambled. You will read each sentence one fragment at a time. Your task is to read as QUICKLY and ACCURATELY as possible. Make sure you understand the meaning of each sentence, just as though you were reading a book or article or blog.

Your task is to:
(1) Press the SPACE BAR to see each new sentence fragment.
(2) Indicate how easy or difficult the sentence was for you to read.

Press the SPACE BAR to continue.

During the experiment, you will read scrambled sentences one fragment at a time.

For each sentence, you will first see a fixation cross in the middle of the screen like this:

+  

Once you see the fixation cross, you may press the SPACE BAR whenever you are ready to start reading the next sentence. Continue pressing the SPACE BAR to advance through each fragment of the sentence.

Please do not read aloud. Read the sentences silently.

It is VERY important to keep your hands close to the computer keys so that you can read as QUICKLY, YET CAREFULLY as possible.

In other words, read as quickly as you can while making sure you understand the MEANING of the sentence.

Press the SPACE BAR to continue.
When you finish reading each sentence, you will be prompted to indicate how difficult or easy it was to read.

The prompt will appear on the screen and ask "How was reading this sentence?"

You must choose a number that best indicates how difficult or easy it was for you to read that sentence. The scale goes from 1 (VERY EASY) to 6 (VERY DIFFICULT).

Please use the full range of responses. Don't just hit easy or difficult for all of them!

Be sure to use the YELLOW number keys to type your answer (not the normal number keys).

You will have 5 seconds to indicate your level of ease/difficulty.

In the next few slides, you will practice this process on a normal English sentence.

Press the SPACE BAR to continue.

If you have any questions, please ask the experimenter now. Otherwise, you are ready to begin.

Remember, your task is to:

1) Read each sentence one fragment at a time, making sure you understand the sentence.

2) Indicate how difficult/easy that sentence was to read.

Press the SPACE BAR to continue.

Recognition Memory Test Instructions

In the next part of this study, you will read twelve sentences.

Importantly, HALF of these sentences are EXACTLY the same as some of the sentences you just read.

The other HALF of these sentences are NOT exactly the same as the sentences you just read.

Your task is to read each sentence (just as you did before) and then indicate whether you have seen that EXACT sentence before.

You will indicate your answer on a scale from 1 (I HAVE SEEN) to 6 (I HAVE NOT SEEN). You will have 5 seconds to indicate each answer.

Please press the SPACE BAR to continue.
Now you are ready to begin. Remember, your task is to:

1) Read each sentence one fragment at a time by pressing the SPACE BAR (just like part one).

2) Indicate on a scale from 1 (I HAVE SEEN) to 6 (I HAVE NOT SEEN) to what extent you think you've seen that EXACT sentence before.

Again, be sure to use the range of responses from 1-6. Don't just say that you've seen them all!

Press the SPACE BAR to begin.

18-Exposure Condition Stimuli - Experimental

A  Earlier today\the farmer\at the market\tomatoes\sold
A  Last month\the waitress\in the restaurant\the dishes\washed
A  Over the weekend\the musician\for charity\a concert\gave
A  Recently\the boy\on the lake\the boat\rowed
A  This morning\the model\at the mall\some perfume\bought
A  Yesterday\the doctor\in the hospital\a surgery\performed
B  Earlier today\the farmer\tomatoes\at the market\sold
B  Last month\the waitress\the dishes\in the restaurant\washed
B  Over the weekend\the musician\a concert\for charity\gave
B  Recently\the boy\the boat\on the lake\rowed
B  This morning\the model\some perfume\at the mall\bought
B  Yesterday\the doctor\a surgery\in the hospital\performed
C  Earlier today\sold\the farmer\at the market\tomatoes
C  Last month\washed\the waitress\in the restaurant\the dishes
C  Over the weekend\gave\the musician\for charity\a concert
C  Recently\rowed\the boy\on the lake\the boat
C  This morning\bought\the model\at the mall\some perfume
C  Yesterday\performed\the doctor\in the hospital\a surgery
18-Exposure Recognition Memory Test Items

A Earlier today the farmer at the market tomatoes sold
B Last month the waitress the dishes in the restaurant washed
C Over the weekend gave the musician for charity a concert
D Recently on the lake the boat the boy rowed
E At the mall this morning the model some perfume bought
F Yesterday performed a surgery the doctor in the hospital

A Recently the boy on the lake the boat rowed
B This morning the model some perfume at the mall bought
C Yesterday performed the doctor in the hospital a surgery
X The farmer earlier today sold at the market tomatoes
Y Last month in the restaurant the waitress the dishes washed
Z Over the weekend gave for charity the musician a concert

48-Exposure Exposure Phase Sentences for the Exp48 Group

Recently the girl at the café her friend met
Over the weekend the athlete at the game his jersey ripped
This morning the model at the mall some perfume bought
Earlier today the farmer at the market tomatoes sold
Over the weekend the musician for charity a concert gave
Some time ago the photographer for his pictures an award won
Recently the policeman for speeding the car stopped
Earlier today the artist for the community a mural painted
This past week the scientist in our class the formula explained
Some time ago the teacher in the book the ideas discussed
Yesterday the doctor in the hospital a surgery performed
Last month the waitress in the restaurant the dishes washed
This morning the mother on television movies watched
Yesterday the girl on the balcony cigarettes smoked
This past week the teacher on the chalkboard the answers wrote
Last month the father on the grill burgers cooked

Recently the girl her friend at the café met
Over the weekend the athlete his jersey at the game ripped
This morning the model some perfume at the mall bought
Earlier today the farmer tomatoes at the market sold
Over the weekend the musician a concert for charity gave
Some time ago the photographer an award for his pictures won
Recently the policeman the car for speeding stopped
Earlier today the artist a mural for the community painted
This past week the scientist explained the formula in our class. Some time ago the teacher discussed the ideas in the book. Yesterday the doctor performed a surgery in the hospital. Last month the waitress washed the dishes in the restaurant. This morning the mother watched movies on television. Yesterday the girl smoked cigarettes on the balcony. This past week the teacher wrote the answers on the chalkboard. Last month the father cooked burgers on the grill.

Recently met the girl at the café. Her friend. Over the weekend ripped the athlete at the game. His jersey. This morning bought the model at the mall. Some perfume. Earlier today sold the farmer at the market. Tomatoes. Over the weekend gave the musician for charity a concert. Some time ago won the photographer for his pictures an award. Recently stopped the policeman for speeding the car. Earlier today painted the artist for the community. A mural. This past week explained the scientist in our class the formula. Some time ago discussed the teacher in the book the ideas. Yesterday performed the doctor in the hospital a surgery. Last month washed the waitress in the restaurant the dishes. This morning watched the mother on television movies. Yesterday smoked the girl on the balcony. Cigarettes. This past week wrote the teacher on the chalkboard the answers. Last month cooked the teacher on the grill.

48-Exposure Condition Recognition Memory Test Sentences
A Earlier today the farmer at the market sold tomatoes.
B Last month the waitress the dishes in the restaurant washed.
C Over the weekend gave the musician for charity a concert.
D Recently at the café her friend the girl met.
E On the chalkboard this past week the teacher the answers wrote.
F Yesterday performed a surgery the doctor in the hospital.
A Over the weekend the athlete at the game his jersey ripped.
B Recently the policeman for speeding the car stopped.
C This morning the model at the mall some perfume bought.
X The photographer some time ago won for his pictures an award.
Y Earlier today for the community the artist a mural painted.
Z Yesterday smoked on the balcony the girl cigarettes.
96 – Exposure Condition Experimental Sentences

Over the weekend\the musician\for charity\a concert\gave
Last month\the model\for her birthday\a party\threw
This morning\the farmer\for his neighbor\some vegetables\picked
Some time ago\the doctor\for his patient\some medicine\prescribed
Some time ago\the photographer\for his pictures\an award\won
This past week\the scientist\in our class\the formula\explained
Recently\the policeman\for speeding\the car\stopped
This morning\the mother\on television\movies\watched
Yesterday\the girl\on the balcony\cigarettes\smoked
Over the weekend\the mother\on the bed\sheets\put
Some time ago\the teacher\in the book\the ideas\discussed
Recently\the girl\at the café\her friend\met
This past week\the teacher\on the chalkboard\the answers\wrote
Earlier today\the artist\for the community\a mural\painted
Recently\the boss\for the company\an accountant\hired
This past week\the musician\at the concert\her violin\played
Over the weekend\the athlete\at the game\his jersey\ripped
Last month\the father\on the grill\burgers\cooked
Yesterday\the doctor\in the hospital\a surgery\performed
Earlier today\the father\in the kitchen\the lightbulb\changed
This morning\the scientist\in the laboratory\the chemicals\examined
Recently\the boy\on the lake\the boat\rowed
This morning\the model\at the mall\some perfume\bought
Earlier today\the farmer\at the market\tomatoes\sold
Last month\the policeman\at the meeting\his concerns\expressed
Over the weekend\the artist\in the museum\her work\unveiled
Yesterday\the boy\in the park\a cellphone\found
Last month\the waitress\in the restaurant\the dishes\washed
Some time ago\the boss\at the store\an employee\fired
This past week\the waitress\on the table\her tip\left
Earlier today\the athlete\on the team\his position\lost
Yesterday\the photographer\at the wedding\some pictures\took

Over the weekend\the musician\a concert\for charity\gave
Last month\the model\a party\for her birthday\threw
This morning\the farmer\some vegetables\for his neighbor\picked
Some time ago\the doctor\some medicine\for his patient\prescribed
Some time ago\the photographer\an award\for his pictures\won
This past week\the scientist\the formula\in our class\explained
Recently\the policeman\the car\for speeding\stopped
This morning\the mother\movies\on television\watched
Yesterday\the girl\cigarettes\on the balcony\smoked

215
Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.

Over the weekend, the mother put sheets on the bed. Some time ago, the teacher discussed the ideas in the book. Recently, the girl met her friend at the café. This past week, the teacher wrote the answers on the chalkboard.

Earlier today, the artist painted a mural for the community. Recently, the boss hired an accountant for the company. This past week, the musician played her violin at the concert.
Last month I cooked the father on the grill burgers
Yesterday performed the doctor in the hospital a surgery
Earlier today changed the father in the kitchen the light bulb
This morning examined the scientist in the laboratory the chemicals
Recently rowed the boy on the lake the boat
This morning bought the model at the mall some perfume
Earlier today sold the farmer at the market tomatoes
Last month expressed the policeman at the meeting his concerns
Over the weekend unveiled the artist in the museum her work
Yesterday found the boy in the park a cellphone
Last month washed the waitress in the restaurant the dishes
Some time ago fired the boss at the store an employee
This past week left the waitress on the table her tip
Earlier today lost the athlete on the team his position
Yesterday took the photographer at the wedding some pictures

96-Exposure Recognition Memory Test Sentences
A Over the weekend the musician for charity a concert gave
B Last month the model a party for her birthday threw
C This morning picked the farmer for his neighbor some vegetables
D Some time ago for his patient some medicine the doctor prescribed
E In our class this past week the scientist the formula explained
F Stopped recently the car the policeman for speeding
A Yesterday the photographer at the wedding some pictures took
B Earlier today the athlete his position on the team lost
C Some time ago fired the boss at the store an employee
X The waitress this past week left on the table her tip
Y Over the weekend in the museum the artist her work unveiled
Z Yesterday found in the park the boy a cell phone

Appendix C
Reaction time measures from the exposure phases of experiment 2

18-EXPOSURE CONDITION

Reaction times were the amount of time (in milliseconds) from the onset of the first syntactic category presented in the self-paced moving window task to the final depressing of the space bar that allowed a participant to finish the sentence. Reaction times (RT) that were +/- 2.5 standard deviations from the mean were discarded, resulting in a loss of 1.1% of RT data from the Experimental group and 1.5% of RT data from the Control group. Table A.1 and Figure A.1 show the mean RTs for Experimental and Control groups in Blocks 1 and 2 (the first 9 and second 9 trials, respectively). As with recognition memory judgments, difference scores were used as a method of normalizing each participant's RT changes. Table A.1 shows the mean RT difference scores, operationalized as mean RT during block 1 minus the mean RT during block 2. Therefore, positive difference scores indicated increasingly faster RTs, while negative difference scores indicated increasingly slower RTs.

Figure A.1. Mean reaction times (in milliseconds) during each trial of the 18-Exposure condition.
To investigate whether participants in the Experimental group or Control group differed in their RT improvement, difference scores between groups were compared. A Mann-Whitney test revealed no significant differences in RT difference scores between Experimental ($M = 476$, $SD = 1445$) and Control participants ($M = 141$, $SD = 401$), $U = 95.0$, $z = -0.44$, $p = .68$. Although the Experimental group’s RT difference score is substantially larger than that of the Control group, it appears that the very large standard deviations may have played a role in the lack of significance. Indeed, the K-S test showed that the Experimental group’s data were non-normally distributed, $D(14) = 0.32$, $p < .001$. Moreover, paired sample t-tests showed no significant differences in RT between Blocks 1 and 2 in the Experimental group, $t(14) = 1.27$, $p = .22$, or the Control group, $t(13) = 1.31$, $p = .21$.

Table A.1. Descriptive statistics for the reading times (in milliseconds) during the training phase in the 18 exposure condition by group and test item type.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>3348</td>
<td>7506</td>
<td>1869</td>
<td>482</td>
<td>2312</td>
<td>4383</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2872</td>
<td>3426</td>
<td>1010</td>
<td>260</td>
<td>2312</td>
<td>3432</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>476</td>
<td>5927</td>
<td>1445</td>
<td>373</td>
<td>-324</td>
<td>1275</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>3851</td>
<td>3169</td>
<td>953</td>
<td>254</td>
<td>3300</td>
<td>4401</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3710</td>
<td>3285</td>
<td>930</td>
<td>248</td>
<td>3172</td>
<td>4247</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>141</td>
<td>1372</td>
<td>401</td>
<td>107</td>
<td>-90</td>
<td>372</td>
</tr>
</tbody>
</table>

Table A.2. Results of tests to check for the normality of distributions of data on reading times in the 18 exposure condition.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Kolmogorov-Smirnov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Skew</td>
<td>SE</td>
<td>Kurtosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>2.15</td>
<td>0.58</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.45</td>
<td>0.58</td>
<td>-0.58</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>3.22</td>
<td>0.58</td>
<td>11.23</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>0.51</td>
<td>0.59</td>
<td>-0.64</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.10</td>
<td>0.59</td>
<td>-0.31</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.85</td>
<td>0.59</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Thus, there appears to be little evidence of a RT speedup. However, independent sample t-tests showed significant differences between Experimental and Control groups in the mean RTs for whole sentences during Block 2, $t(27) = 2.31, p = .02, d = 0.89$, but not during Block 1, $t(27) = 0.90, p = .37$ (see Figure A.2 below). Thus, it is possible that Experimental and Control participants may have truly differed in some respects in their RTs. To further investigate this possibility, a mixed ANOVA was conducted with Group as a between-subjects factor (2 levels: Experimental, Control) and Block as a within-subjects factor (2 levels: Block 1 and 2). The ANOVA revealed no main effects of Group, $F(1, 27) = 2.48, p = .13$, Block, $F(1, 27) = 2.37, p = .14$, and no Group*Block interaction, $F(1, 27) = 0.69, p = .41$. Thus, there is little evidence overall that participants in either group got faster at reading sentences, although the Experimental group’s reading times during Block 2 were significantly faster than those of Controls.

Figure A.2. Mean reading times for whole sentences during the exposure phase in the 18 exposure condition separated by Block (2 blocks, 9 trials each). Error bars represent standard deviations.
48-EXPOSURE CONDITION

Recall that RTs were the amount of time (in milliseconds) from the onset of the first syntactic category presented in the self-paced moving window task to the final depressing of the space bar that allowed a participant to finish the sentence. RTs that were +/- 2.5 standard deviations from the mean were discarded, resulting in a loss of 1.7% of RT data from the Experimental group and 1.2% of RT data from the Control group. Figure A.3 shows the mean trial-by-trial reading times for the Experimental and Control groups across the training phase.

Table A.3. Descriptive statistics for the reading times (in milliseconds) during the training phase in the 48 exposure condition by group and test item type. RTs for the below data were normally distributed, all K-S ps > .07.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>3550</td>
<td>2823</td>
<td>899</td>
<td>240</td>
<td>3030</td>
<td>4070</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3011</td>
<td>2406</td>
<td>739</td>
<td>197</td>
<td>2584</td>
<td>3438</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3080</td>
<td>3799</td>
<td>987</td>
<td>263</td>
<td>2510</td>
<td>3650</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2831</td>
<td>2775</td>
<td>693</td>
<td>185</td>
<td>2431</td>
<td>3232</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>324</td>
<td>1332</td>
<td>416</td>
<td>111</td>
<td>84</td>
<td>565</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>3876</td>
<td>5177</td>
<td>1434</td>
<td>397</td>
<td>3009</td>
<td>4743</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3733</td>
<td>6168</td>
<td>1636</td>
<td>453</td>
<td>2744</td>
<td>4721</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3791</td>
<td>5661</td>
<td>1669</td>
<td>462</td>
<td>2783</td>
<td>4800</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3703</td>
<td>5331</td>
<td>1523</td>
<td>422</td>
<td>2782</td>
<td>4623</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>57</td>
<td>1709</td>
<td>530</td>
<td>147</td>
<td>263</td>
<td>377</td>
</tr>
</tbody>
</table>

In order to determine whether there were overall differences in RT speedup between groups, difference scores were computed for each participant as the mean RT in the first 24 trials of training minus the mean RT in the last 24 training trials. An independent-samples t-test revealed no significant difference in RT difference scores between the Experimental group ($M = 324$, $SD = 416$) and the Control group ($M = 57$, $SD = 530$). However, this finding should be interpreted with caution, since the RT difference scores are substantially different and the standard deviations suitably large to obscure a speedup effect in the Experimental group.
Therefore, further analyses were conducted to see if RT speedup differences between groups were evidenced in the individual training blocks. To further investigate RT differences, a 2x4 mixed ANOVA with Group (Experimental, Control) as between-subjects factor and Block (1, 2, 3, 4) as within-subjects factor was performed on participants’ mean reaction times in each block (a Block consisted of the presentation of 12 sentences). The ANOVA revealed a main effect of Block, $F(3, 75) = 6.96, p < .001, \eta^2_p = .23$, but not of Group, $p = .16$. The Group*Block interaction approached significance, $F(3, 75) = 2.61, p = .058, \eta^2_p = .09$. Moreover, planned post hoc tests comparing mean RTs in each block between groups revealed no significant differences in mean RT between groups in Blocks 1-3, $ps > .15$, and differences approached significance in Block 4, $t(25) = 1.93, p = .06$. Thus, despite numerical evidence of superior speedup in the
Experimental group’s difference score, the effect is not robust enough to result in a statistically-reliable speedup.

Figure A.4. Mean reading times (in ms) in each block of the exposure phase for Experiment and Control groups. Error bars represent standard deviations.

### 96-EXPOSURE CONDITION

RTs that were +/- 2.5 standard deviations from the mean were discarded, resulting in a loss of 1.1% of RT data from the Experimental group and 1.2% of RT data from the Control group. A K-S test was used to assess the normality of the distribution of RTs and showed that the RTs were normally distributed within blocks and that the RT difference scores were also normally distributed, all \( ps > .15 \).
In order to determine whether there were overall differences in RT speedup between groups, difference scores were computed for as the mean RT in the first 24 trials of training minus the mean RT in the last 24 training trials. An independent-samples t-test revealed no significant differences in mean difference scores between the Experimental group ($M = 314, SD = 655$) and Control group ($M = 224, SD = 515$), $t(24) = .38$, $p = .70$. 

Table A.4. Mean reading time latencies and difference scores for Experimental and Control participants across the exposure phase blocks of the E96 condition. Each block consists of 12 trials.

<table>
<thead>
<tr>
<th>Group</th>
<th>Block</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bound</td>
</tr>
<tr>
<td>Experimental</td>
<td>1</td>
<td>3939</td>
<td>4653</td>
<td>1364</td>
<td>378</td>
<td>3115</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3560</td>
<td>4762</td>
<td>1272</td>
<td>352</td>
<td>2791</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3635</td>
<td>4516</td>
<td>1298</td>
<td>360</td>
<td>2850</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3541</td>
<td>3939</td>
<td>1084</td>
<td>300</td>
<td>2887</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3408</td>
<td>3924</td>
<td>1086</td>
<td>301</td>
<td>2751</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3583</td>
<td>4012</td>
<td>1204</td>
<td>333</td>
<td>2855</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3472</td>
<td>3323</td>
<td>1036</td>
<td>287</td>
<td>2846</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3398</td>
<td>3216</td>
<td>1021</td>
<td>283</td>
<td>2781</td>
</tr>
<tr>
<td>Difference</td>
<td>1</td>
<td>314</td>
<td>2131</td>
<td>655</td>
<td>181</td>
<td>-81</td>
</tr>
<tr>
<td>Control</td>
<td>2</td>
<td>3631</td>
<td>4664</td>
<td>1412</td>
<td>391</td>
<td>2778</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3073</td>
<td>3521</td>
<td>1076</td>
<td>298</td>
<td>2422</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3094</td>
<td>4320</td>
<td>1301</td>
<td>361</td>
<td>2308</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3068</td>
<td>4504</td>
<td>1206</td>
<td>334</td>
<td>2339</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2996</td>
<td>4242</td>
<td>1205</td>
<td>334</td>
<td>2267</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3039</td>
<td>4047</td>
<td>1164</td>
<td>323</td>
<td>2335</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3139</td>
<td>5108</td>
<td>1327</td>
<td>368</td>
<td>2336</td>
</tr>
<tr>
<td>Difference</td>
<td>2</td>
<td>224</td>
<td>2202</td>
<td>515</td>
<td>142</td>
<td>-86</td>
</tr>
</tbody>
</table>
Further analyses were conducted to see if RT speedup differences between groups were evidence in the individual training blocks. To further investigate RT differences, a 2x8 mixed ANOVA
with Group (Experimental, Control) as between-subjects factor and Block (1-8) as within-subjects factor was performed on participants’ mean reaction times in each block (a Block consisted of the presentation of 12 sentences). The ANOVA (with Greenhouse-Geisser correction for violation of sphericity) revealed no significant effect of Block, $F(3.99, 95.92) = 1.13, p = .34$, no Block*Group interaction, $F(7.99, 95.92) = .29, p = .99$, and no effect of Group, $F(2, 24) = .45, p = .64$. Thus, there was no evidence of a significant difference in RT speedup between groups, suggesting that any actual RT gains are probably due to practice effects.
REFERENCES


Hamrick, P., & Sachs, R. (in prep). Trained control groups in implicit learning research in SLA.


measures. Paper presented at the conference for the American Association for Applied Linguistics. Dallas, TX, USA.


Williams, J. N. (2010). Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning, 60*, 221-244.


