Bridging between on-line linguistic adaptation and long-term language learning

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Marius Janciauskas

October 2017

Table of Contents

LIST OF ABBREVIATIONS

AGL	Artificial grammar learning
ANOUN	Animate noun category
DO	Double object dative
DVERBD	Dative verb biased towards double object dative structure
DVERBP	Dative verb biased towards prepositional object dative structure
IN	Intransitive structure
IVERB	Intransitive verb
INOUN	Inanimate noun category
LAMOLL	Linguistic Adaptation Mechanism of Language Learning
NP	Noun phrase
PD	Prepositional dative
PP	Prepositional phrase
PREP	Preposition
RT	Reaction time
RSVP	Rapid serial visual representation task
SRT	Serial reaction time task
SRN	Simple recurrent network
TR	Transitive structure
TVERB	Transitive verb

Abstract

Linguistic adaptation is a set of phenomena where language representations change in adults in response to linguistic input. Some of these adaptation effects are persistent and have been argued to be due to learning mechanisms that are similar to those used in language acquisition (Dell & Chang, 2013). Theories of language acquisition often assume that the creation of language-specific categories and structures involves bigger changes in linguistic representations than those required to support linguistic adaptation and it's not clear how their proposed mechanisms extend to explaining learning effects in adults. This thesis will examine the extent to which common mechanisms could explain both language acquisition and linguistic adaptation. In the first part of the thesis, we use a non-linguistic artificial grammar learning task to teach participants a simplified language over the course of an experiment. During the experiment, we examine the extent to which sentence-grain linguistic adaptation effects are related to the experiment-grain acquisition of the structural rules of the language. By linking these processes we provide behavioral evidence that short-term effects that occur at the level of sentence processing and longer-term effects that are characterized by structure acquisition are supported by a common mechanism. The second part of the thesis attempts to link the learning processes that take place in the artificial grammar learning tasks to the real-world language learning that takes place over the course of years and that result in long-term structure knowledge To do so, we use a computational model that is able to simulate the sentence/experiment grain changes to model how learners acquire a second language (L2). Since the data from L2 learners captures the changes that arise from years of real world learning, the model provides a link between sentence/experiment grain learning and year-grain learning. The final chapter links these two sections by examining how some of the behaviours observed in L2 speakers could be modeled within our artificial grammar learning paradigm.

Chapter 1 provides an overview of the complexities that child language acquisition studies must explain and introduces the linguistic adaptation phenomena in adults highlighting the need to incorporate these processes into the theories of language learning.

It will then describe the existing language acquisition theories focusing on whether they could explain linguistic adaptation effects like verb bias and abstract structural priming. The chapter will then proceed by describing a prediction-based statistical learning mechanism that is based on the assumptions of the connectionist model of language production, the Dual Path model (Chang, 2002). Referred to as Linguistic Adaptation Mechanism of Language Learning (LAMOLL), it will detail a potential unified account of language structure acquisition and linguistic adaptation in adults.

To support this account, Chapter 2 will present 4 studies that use a non-linguistic artificial grammar-learning (AGL) task to investigate if grammar acquisition and linguistic adaptations effects like structural priming could be elicited in a single grammar learning task. The studies will also test the prediction-based nature of this learning mechanism.

Chapter 3 will shift attention to demonstrating that the same learning mechanism is active in children and adults. Second language studies show that people who start learning an L2 at a later age do not generally learn it to the same degree as native language learners do. This is often used to support the claim that language learning relies on language-specific learning mechanisms that are not active in adults. Such claims challenge the assumptions that the same mechanism that is responsible for language acquisition could support linguistic adaptation effects in adults. In this chapter we will explore the lack of the effect of years of language input on adult grammar knowledge reported in these studies employing a variety of methods ranging from corpus analysis and data reanalysis to modeling the results using the Dual Path model that motivated the use of AGL task to study language structure acquisition and adaptation effects in Chapter 2.

Chapter 4 will report two experiments that will employ the AGL task developed in Chapter 2 to explore behaviorally the L2 effects observed in Chapter 3. This will test how the new AGL task generalized to new situations and its ability to explain linguistic effect as a method to study domain-general learning effects in language.

Finally, Chapter 5 will summarize the findings from all studies discussing their implications and will provide some directions for the future studies.

Acknowledgements

I would like to thank my primary supervisor, Dr. Franklin Chang, for the guidance, patience and his time invested in my supervision. It has been a great pleasure exploring the depths of language research under your guidance, testing ideas that ranged from bad and mad to wild and dazzling. I could not have wished for a better supervisor and I am truly lucky to have studied under your guidance.

My special thanks go to Prof Caroline Rowland who got me hooked on language research while I was still a second-year undergrad student. Many years later I'm still wondering what structural priming is, so thank you for introducing me to the invigorating world of psycholinguistics and for supporting me as a secondary supervisor.

I would like to thank all members of our research group, including Prof Elena Lieven, Prof Julian Pine, Prof Ben Ambridge, Dr. Daniel Freudenthal, Dr. Colin Bannard and others for the invigorating discussions, shared knowledge, and support behind the scenes.

I would also like to thank the following people who made my Ph.D. more eventful: Amy Bidgood for many chats and laughter, Katie Twomey for sharing supervisor survival tips and countless hours of explaining silly things that where beyond my competence at the time, Pawel Jedras for cruel jokes, shared stress and encouragement, and of course two other special members of the MMM team (Michelle Peter, Maxine Sharps) for the great times we spent procrastinating.

Finally, I would like to thank Economic and Social Research Council for funding this Ph.D.

Rationale for submitting the thesis in an alternative format

This thesis has been submitted in the Alternative Paper Format, which consists of chapters that have been presented in a format that is suitable for publication in a peer-reviewed journal. This alternative format meets the high standard expected from a traditional thesis in all aspects and the motivation behind it is as follows.

The chapters in this thesis are either published or are being prepared for publication. The reason for adopting this format is that peer-review process is an integral part of scientific research and reviewers' feedback ensures that the research presented as a part of this thesis is current and relevant. It also allows sharing the research finding with the peers in the field enriching the existing literature and stimulating the discussions around the topic. At the same time, it ensures better career prospect and recognition in the field.

Another reason for adopting this format is that the thesis contains a multifaceted argument that rests on the evidence drawn from the studies that are not directly related to each other or follow up one another naturally. The current format helps to improve the flow and coherence of the thesis. As in the traditional format, this work contains a general introduction that describes the details of the thesis and identifies different levels of evidence required to support its argument. Each chapter is then presented in a paper format and consists of its own introduction, experiments, discussion and conclusion that address the hypotheses specific to that paper. Each chapter is also preceded by a short section, connecting it to the overall structure of the thesis. Finally, the discussion summarises the findings from each study that are relevant to the argument, explaining how the studies fit into the wider context of the thesis. All references are listed together at the end of the thesis.

As of the thesis submission date, one paper (Chapter 2) is being revised for resubmission, and one paper (Chapter 3) is published in Cognitive Science. While the papers have been written with co-authors fulfilling advisory roles, I was responsible for literature research, design, piloting and testing, data analysis, connectionist design and modeling, writing and submitting the papers.

Chapter 1. Introduction to language acquisition and linguistic adaptation

1. Relationship between linguistic adaptation and language acquisition

The mechanisms that support language learning have puzzled cognitive scientists for decades. While children who receive appropriate language input learn language structure quite effortlessly, the same could not be said about adult second language learners and computers that fail to acquire language knowledge that is as general and natural as the knowledge in children. Likewise, adult language studies show that language users continuously update their language representations in response to language input and these learning effects seem to be less rigorous and qualitatively different from the fast learning processes that children undergo to learn those representations. This thesis will examine a mechanism that could provide a link between child language acquisition and adult linguistic adaptation.

The complexity of explaining the mechanism that supports language learning comes from a large number of rules and constraints that support the processing of even simple sentences. This can be illustrated with the sentence *The boy gives a book to the girls.* The ability to produce and understand such sentences requires the knowledge that the words *boy* and *girls* can be substituted to express different meaning but words *boy* and *gives* can't, and this knowledge comes from learning that *boy* and *girls* are part of the same syntactic category of noun. In addition to the category knowledge, we know that the noun *girls* is the plural form of the noun *girl*. Furthermore, there are various regularities between words that, for example, require the use of the verb *gives* in the third person singular form because the subject *boy* is a singular noun. Language users must also know how to order categories in a sentence. For example, nouns like *boy*, *girls* and *book* are preceded by the articles *a* or *the* and their use depends on the form of the noun (e.g. *a book* but not *a books*).

In addition to these word level constraints, there are also higher-level rules. For example, the meaning of the sentence *The boy gives a book to the girls* can be conveyed by another sentence *The boy gives the girls a book*. The first structure is called the prepositional dative (PD) structure, since it has a prepositional phrase (*to* the girl), while the second is called the double object (DO) structure, as it has two object noun phrases after the verb. The choice between these structures is often used to study the knowledge that supports sentence production. Also, many verbs (e.g. *give*) can alternate between these two structures but some verbs can only appear in one or the other structure. For example, while similar to the verb *give* in its meaning, the verb *donate* can only occur in the PD structure, but not the DO structure (* *The boy donates the girl the book*). The structures that are acceptable with a verb are part of their subcategorisation. However, in addition to these more binary constraints, verbs are also biased towards particular structures (e.g., *give* occurs more frequently in the double object than the prepositional dative).

The examples above illustrate just some of the constraints that guide grammatical sentence construction and explaining how children acquire adult-like categories, inflection, sentence structures and structural alternations from the input has been the focus of language acquisition theories. However, recent adult language processing studies have also shown that language learning does not end with the attainment of adult-like knowledge. There is evidence that linguistic knowledge continuously changes throughout life and this process is referred to as linguistic adaptation. One such example concerns verb bias. For instance, language speakers use verb *give* in the double object dative more frequently than in the prepositional dative structure (Campbell & Tomasello, 2001; Gries & Stefanowitsch, 2004; Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989). This effect can be measured both in production studies (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Wilson & Garnsey, 2009) and in comprehension studies where verb bias creates expectation of the structure with which it is associated more strongly (Garsney et al., 1997; Osterhout, Holcomb, & Swinney, 1994; Trueswell, Tanenhaus, & Kello, 1993). Although these biases are consistent in adults, they also seem to change in response to the input that people hear. For example, Coyle & Kaschak (2008) manipulated the verb bias preferences for the verbs *give*, *loan*, *send* and *hand*. Participants in their study were selected based on their strong preference to use each verb with one of the structures. In the training phase, participants were asked to complete 20 sentence fragments like (1) and (2), that were designed to elicit either DO construction or PD construction respectively. For example, sentence fragment (1) contained an animate noun *student* after the verb, which made people more likely to continue the sentence using a DO structure. Fragment (2), on the other hand, had an inanimate noun *book* after the verb, which made people more likely to use a PD structure to complete the sentence.

(1) The teacher sent the student…

(2) The man handed the book…

(3) The mechanic sent…

(4) The professor handed…

After the training phase, the participants had to complete sentence fragments that ended with a verb only (3) (4). They found that biasing a verb toward a particular construction in the training phase increased the probability that participants would use that construction when completing target stems involving that verb in the priming phase regardless of the participant's initial preferences with those verbs. For example, *send* is a PD-biased verb but exposing participants to the same verb in DO construction made them more likely to complete the sentence with this verb using DO structure. This shows that verb-structure relationships are not fixed but that they are adapting to the regularities in the input.

Similar adaptation effects are also evident at the abstract level of language structure. One such phenomenon is known as structural priming (Bock, 1986). For example, in one of her experiments, the participants had to repeat sentences produced by the experimenter that had either DO or PD structure like (5) or (6). After such a prime sentence, they were asked to describe a semantically unrelated picture that, for example, depicted a police officer, a driver, and a ticket. They could describe it using either DO or PD structure sentence like (7) or (8) but the author found they were more likely to describe the picture using the same sentence structure that they had heard previously.

(5) A rock climber sold an undercover agent some cocaine

(6) A rock climber sold some cocaine to an undercover agent

(7) A police officer issued the driver a ticket

(8) A police officer issued a ticket to the driver

These changes are thought to involve abstract structure because priming is not changed by manipulation of the overlapping function words like prepositions (e.g. *to* an undercover agent vs. *to* the driver) that could bias people to produce sentences with the same prepositional phrase. For example, Bock (1989) varied the preposition in the prime sentence (e.g. *to* her boss vs. *for* her boss) but found no differences between prepositions in terms of priming rates, suggesting that priming involved abstract structure independent from the repetition of specific words. Additional evidence for the abstract nature of these structures comes from studies that manipulate the thematic roles in the events. Bock and Loebell (1990) found that prepositional locative sentences like *The wealthy widow drove the Mercedes to the church* made people likely to describe target picture using prepositional dative sentence to the same degree as prepositional dative sentence like *The wealthy widow gave the Mercedes to the church*, even though they differ in event role of the preposition. That is, in the preposition locative, the prepositional phrase encodes the location, while in the dative it encodes the recipient. However, at the structural level both sentences have the same abstract structure that could be expressed as NP VP [V NP PP[P NP]], where NP is a noun phrase (e.g. *the wealthy widow*), VP is a verb phrase that consists of a verb (V; e.g. *drove*), a noun phrase (e.g. the Mercedes) and a prepositional phrase (PP), that consists of a preposition (P; e.g. *to* or *for*) and a noun phrase (*the church*). So while priming can be influenced by the overlap in content words like verbs or nouns (Pickering & Branigan, 1998; Scheepers, Christoph, Raffray, Claudine and Myachykov, Andriy, 2017), there is a purely abstract component that exists independent of these additional effects.

Importantly, structural priming effects persist over time. If people are exposed to a prime sentence, its effects are evident even if as many as 10 structurally unrelated sentenced are presented between a prime and a target (Bock and Griffin 2000; Bock, Dell, Chang & Onishi. 2007). Recently, Branigan and Messenger (2016) have found that priming effects last both in children and adults a week after the priming session. This means that the effect of priming persists in the language processing system, and that suggests that learning has taken place.

In summary, speakers rely on a wide range of linguistic knowledge to process sentences like PD and DO and this linguistic knowledge is learned in childhood during language acquisition. Linguistic adaptation effects suggest that language representations continue to change in adults even after a single instance of processing a sentence and we will refer to the changes at this level as the sentence-grain level changes. Language acquisition, on the other hand, takes place over many years and involves the creation of new categories and structures. It is not clear, whether sentence-grain changes can be explained assuming the same mechanisms that are needed for explaining year-grain language learning. To examine this, we will first review the existing theories of language acquisition. We will then proceed by discussing an account that could potentially link language acquisition and linguistic adaptation assuming a common learning mechanism.

2. Language acquisition theories

Language acquisition has been generally approached from two very different perspectives, broadly known as generativist and constructivist. They take very different stances on the nature of language representations and the mechanism that support their acquisition. Each approach will be described separately focusing on the mechanisms that support the learning of syntactic categories, inflection, abstract sentence structures, and verb bias. These domains will be important in the studies throughout the thesis, so it is important to review the existing work in order to understand the relationship between language acquisition and linguistic adaptation phenomena.

2.1. Generativist approach to language acquisition

Generativist theories of language acquisition emphasize the role of innate knowledge in language learning. This knowledge, formally known as Universal Grammar (Chomsky, 1975), is a set of language-specific parameters and constraints on the type of rules that support different languages. They help ensure that children quickly identify the rules that are needed for a particular language, while at the same time ruling out a large number of possible rules which don't exist in any language (Chomsky, 1965; Pinker, 1984; Valian, 1986). Since different languages have different grammar rules (e.g. inflection rules, word order) many generativists (e.g. Yang, 2002) assume the existence of multiple language systems that have language-specific rules. The particular system is activated with the help of mental parameters that children set based on the evidence for or against certain systems in the child's input. In other words, under the generative approach children are born with the knowledge about language structure and the sole purpose of learning is to activate the language-specific knowledge appropriate for the specific language that the child is learning (Chomsky, 1965; Pinker, 1984; Valian, 1986).

2.1.1. Inflection

To explain inflection knowledge, generativists postulate that children are born with the knowledge of abstract categories like verbs and tense inflection (e.g. -s,), and the rules that allow combining the two (Pinker, 1984; Harris & Wexler, 1996; Hoekstra & Hyams, 1998; Wexler, 1998; Legate & Yang, 2007). In other words, children do not need to learn different instances of how word stems (e.g. give, throw, take) combine with inflections (e.g. –s) but instead, they know innately that the inflection can be combined with any members of the verb category.

Since not all languages use inflection, children have special inflectional parameters that help them learn if the language marks words for tense like, for example, English or not. This parameter helps the system make more fine-grained distinctions in how it applies inflectional forms. Evidence for this approach comes from corpus studies of Spanish, French and English languages that showed that children stop making errors (e.g. omitting –s from 3rd person singular verbs) in these languages at different ages because of the availability of evidence for or against tense marking system in these languages (Legate & Yang, 2007). Critically, these counts were made based on the syntactic labeling of these forms, not using the lexical/morphological forms, and that supports the view that children are using abstract categories as they collect statistical knowledge.

2.1.2. Syntactic categories

Syntactic categories are also argued to be a part of Universal Grammar and they are present in language learners from birth (Pinker, 1989). In other words, a child is born with the knowledge of categories like determiner, verb, noun, and preposition. To identify which words belong to which syntactic category, Pinker (1989) proposed that children's innate knowledge also consists of semantic categories and a set of linking rules that link the two. Semantic categories (e.g. person/thing, action/change of state) capture the meaning of the words and, unlike syntactic categories, are observable in the real world. In other words, the child may not know that 'boy' is a noun but he/she can see that 'boy' is a person. Once the semantic category of the word is identified, they can link them to appropriate syntactic categories automatically (e.g. person/thing -> noun, action/change of state -> verb). Indirect evidence for this account comes from corpora studies showing that there is a very high correspondence between semantic and syntactic categories (Rondal and Cession,

1990) and models have been developed which use this mechanism to support language acquisition (Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017).

2.1.3. Ordering categories

Once children acquire categories, the way they order them (e.g. determiner before noun) is also guided by the innate knowledge of phrase structure rules that dictate how different categories relate to each other. One such theory is X-bar theory described by Chomsky (1970). It claims that phrases are extensions of a head category of that phrase (e.g. the head of a noun phrase is a noun). For example, children must learn to use *the* before *girls* but they also need to know that the determiner is not always adjacent to the noun (e.g. *the* big *girls*). Children know this innately because their phrase structure rules dictate that there are two additional category types: specifiers and adjuncts. In the phrase *the big girls* determiner “the” is the specifier and adjective “big” is the adjunct. Their phrase structure rules dictate that adjuncts can combine with nouns (e.g. big girls) but they can’t combine with noun phrases (e.g. big the girls).

Evidence for such phrase structure rules come from Bloom (1990) who studied children’s speech at the age of 1-2 years. The X-bar theory predicts that since names (e.g. Fred) and pronouns (e.g. they) are noun phrases (e.g. *they* can replace *the girls*), children should automatically avoid attaching adjectives to either pronouns or names (e.g. big Fred, big they) even if they can appear in similar sentence position in the input (e.g. Girls like books; They like books, Fred likes books). However, both nouns and noun phrases can occur with predicative adjectives (e.g. girls are big, they are big, Fred is big). The study found that from the earliest age children have systematic bias not to use pronominal adjectives with names and pronouns but don’t show the same bias for predicative adjectives. They say things like *big dog* frequently but don't say *big Fred* or *big he*. Such biases suggest that children appear to have the concept of phrases and phrase rules that guide the ordering of syntactic categories.

2.1.4. Syntactic structures

The process of combining phrases into sentences like PD and DO in generativist view is also supported by the innate knowledge of word order rules. The child needs to activate the right set of parameters appropriate for that language (Chomsky 1981). These parameters are not specific to the particular structures like DO and PD but apply to classes of constructions that share certain structural properties. The goal of learning is to recognize the properties of the input that would allow switching the right parameters reliably. Consequently, generative theories predict that children should acquire syntactic structures early on in the development. Support for this comes from a study by Snyder & Stromswold (1997), who found that children first learn DO structure at around 1;8 to 2;11 and then PD structure at around 2;0 and 3;4. There was no significant correlation between the frequency of PD in children's input and the age at which PD was acquired, showing that they did not build those structures based on their experience with those structures. Instead, the authors found that DO acquisition was correlated with the acquisition of structurally related constructions like Verb (V)- Noun Phrase (NP) - Particle (e.g. the boys put (V) the shoes (NP) on (Particle)) and PD was correlated with the acquisition of constructions like V-Particle-NP (e.g. the boys put on the shoes). According to the authors, this reflected the fact that the two constructions dependent on discovering two different properties of the language before the right parameters were set to allow the formation of both constructions and thus supports the parametric account of language acquisition.

2.1.5. Verb-structure links

Finally, an important type of evidence that supports generative accounts is over-generalizations. These are the cases when children produce an utterance that is ungrammatical in the adult language and the production of these utterances could not be explained by input-based learning. For example, Bowerman (1987) noted that children say ungrammatical sentences like "you put me just bread and butter". The verb *put* cannot appear in DO structure and hence children never hear it used in this construction. Thus, the fact that they can combine them in novel ways suggests that their representations are not completely dependent on the verb-structure pairs in the input.

Pinker (1984) argues that children make these errors because their structural knowledge is guided by the knowledge of broad range rules that capture the meaning associated with different structures. For example, DO structure is said to denote a change of possession (CAUSE-TO-HAVE). That is '*the boy gives the girls a book'* implies that the boy causes the girls to have a book. PD structure, on the other hand, is said to denote a change of location (CAUSE-TO-GO) and the sentence '*the boy gives a book to the girls'* implies that the boy caused a book to move to the girls. A broad range rule allows children to use the verbs that satisfy cause-to-have meaning in an alternative structure and vice versa. So if a child hears a grammatical utterance like '*You put bread and butter on my plate*', they can map it to the CAUSE-TO-MOVE meaning. However, the broad range rule also allows changing this to the CAUSE-TO-HAVE meaning and this allows them to generate a DO structure with the verb put '*You put me the bread and butter*'. This can explain why children make novel over-generalizations. Support for this rule is provided by Conway and Demuth (2007) who found that when 3-year old children were presented with novel verbs in either DO or

PD sentence (*you pilked the cup to Toby* vs. *you pilked Toby the cup*), they readily generalized them to produce alternating structure in their own speech without having heard verb *pilk* in an alternative structure.

Pinker (1984) argued that children gradually learn to avoid overgeneralizations by learning narrow range rules that put constraints on how verbs are used in different structures. These constraints are based on the semantic categories that capture fine distinctions in the meanings of verbs like *pull*, *deny*, *carry*, *put* in terms of the manner of motion, type of communication, direction and other characteristics. So while the verbs may meet the requirements of broad range rules, its behavior with different constructions is constrained by the semantic category that it belongs to. Verb *put* is part of a semantic category that specifies the result of an event and verbs of this class appear in PD structure only. Children must learn that *put* belongs to this category and until they learn this narrow range rule, they will continue to make overgeneralization errors. Support for this account comes from Ambridge et al., (2012) who observed that semantic constraints like broad range and narrow range rules were one of the significant predictors of children's errors with dative sentences.

2.1.6. Summary: Generativist approach to language acquisition

In sum, generative theories argue that language structure learning is supported by innate, language-specific learning mechanism that involves the activation of pre-existing language knowledge in the form of inflection rules, abstract syntactic categories, phrase structure, and broad range rules. To reach adult-like language abilities, children must map their language input onto the pre-existing representations to activate the appropriate language system that imposes the constraints on language use. This allows children to reach adult-like language performance relatively quickly and effortlessly.

2.2. Constructivist approach to language acquisition

Constructivist theories (also knows as usage-based, input-based) stand in opposition to the generativist approach. These theories assume that children learn language structure by storing lexically-specific instances of their language and then gradually abstracting over them to form more abstract language representations (Tomasello, 2000; Lieven & Tomasello, 2008). Importantly, language knowledge in these theories is emergent in nature and is formed slowly by extracting the regularities directly from the language input. Instead of assuming innate language-specific learning mechanism, these theories argue that learners do so using general-purpose learning mechanisms such as statistical learning (Saffran & Thiessen, 2007).

2.2.1. Inflection

To learn inflection rules, children initially learn whole words like give/gives, jump/jumps, show/shows. Eventually, they schematize across the inflected verbs to form slot-and-frame patterns like X+s, where X can be any verb in the child's repertoire. They can apply this frame to new verbs, thereby acting as abstract inflection rule. Children learn these rules separately for different tenses, such as adding *-ing* to continuous tense verbs (jump-ing) or *–ed* to past tense verbs (jump-ed). Evidence for this mechanism comes from a study by Pine et al. (1998). The authors observed that there was hardly any overlap between children's use of the 3sg present tense *–s*, present progressive *–ing* and regular past tense *–ed* inflections. For instance, there was no child who used both past tense inflection and a 3sg present inflection with the same verb, suggesting that children are not operating with abstract grammatical categories or rules but with categories that are more limited in scope and that they learn each rule from the input.

2.2.2. Categories

In constructivist views, terms like noun and verb are just mnemonics used to describe category-like behaviour. According to this approach, category knowledge comes from people's ability to extract distributional regularities with which words occur in language input (Monaghan, Chater, Christiansen 2005; Monaghan, Christiansen, Chater, 2007; St Clair, Monaghan, & Christiansen, 2010). For example, according to the slot-and-frame approach (e.g. Tomasello, 2003), children track similarities with which different strings of words occur and schematize across them to derive categories of words that occur in the same context. For instance, a child may hear strings of words like *a boy, a girl, a book* and form a lexically specific slot-and-frame schema *a* [X], where X stands for any word that comes after the determiner *a*. A child may also notice strings like *the pen*, *the car, the man* and form a different lexical specific slot-and-frame schema *the* [Y], where Y stands for any words that follow the determiner *the*. Eventually, children abstract their knowledge across different schemas to realize the commonalities between X and Y words and conclude that words that appear with *a* can also be used with *the*, resembling noun category-like knowledge.

A similar mechanism, referred to as frequent frames, was described by Mintz (2003). He proposed that children group together words that appear in frequent frames. In his theory, a frame is defined as two jointly occurring words with one word intervening between them (e.g. a X to). Corpus analysis showed that in most cases such frames were made up from NOUNS (e.g. a book to…), which means that children could exploit distributional information to identify which new words may share the category with the existing known words. The fact that people can indeed use frequent frame-like contexts to infer category membership was also demonstrated experimentally in adults (Mintz, 2002) and infants (Mintz, 2006).

Yet another mechanism was described by Hunt and Aslin (2010). In an artificial language study, the authors demonstrated that people used statistical learning abilities to extract category membership by exploiting distributional patterns that members of different categories share with the members from the preceding and following categories. Such distributional patterns are captured by two statistics: forward transitional probabilities and backward transitional probabilities. Forward transitional probability is the probability that the element will be X given that the current element is Y (e.g. probability of the word *boy* given *the*). Backwards probability is the probability that the previous element was X given the current element Y (e.g. probability that *boy* was preceded by *the*). Participants are sensitive to these statistics and learn that words that have a high degree of transitional probability before or after a group of common words belong to the same category. Their study showed that participants could readily generalize this knowledge to new members of the category even if that specific transition was not observed in the input. In natural language terms, this resembles an instance where people accept a never-before-seen transition between *a* and *boy* because both *boy* and *girl* have a high transitional probability following the word *the* and because the word *girl* also has high transitional probability following *a.*

2.2.3. Ordering categories

In constructivist view, children learn many aspects of structure like syntactic categories and the relationships between different categories (e.g. DET NOUN -> the boy) simultaneously. Like with category formation and inflection, children learn to order categories through abstraction over slot-and-frame patterns (Ambridge & Lieven, 2015). That is, while children learn that words that occur after *a* or *the* belong to the same category of words, they also learn the association between determiner category and noun category, which results in rule-like knowledge that nouns occur after determiners.

An alternative mechanism that follows the statistical learning framework postulates that children learn the relationship between different categories by extracting probabilistic relationships between them. That is, like in category learning, people use transitional probabilities between categories to learn phrase structure-like relationships. For example, Saffran (2001; 2003) exposed people to an artificial language that consisted of three simple phrases made up of combinations of either one or two categories of nonsense words. After exposing participants to auditory sequences of this language, the authors observed that in a grammaticality judgment test people treated sequences that had higher transitional probabilities between categories as more grammatical. Since transitional probabilities between different categories were higher within phrases than between phrases, their knowledge resembled phrase structure that was induced from distributional information in the input. Importantly, these abilities were demonstrated both in children and adults.

2.2.4. Syntactic structure

Constructivist accounts do not offer a clear view on what mechanism could support the learning of constructions like PD and DO. In their view, just like categories and basic word order, structures are built from slot-and-frame schemas observed in the input. It is likely that children first use them in transitive constructions, so it may be the case that they aid the abstraction of dative knowledge (Campbell & Tomasello, 2001). That is, children may derive simpler slot-and-frame patterns that they later combine into more complex constructions before arriving at the fully abstract knowledge of dative structure. For instance, they may use sentences like *I'm pushing the toy* to derive *I'm [X]ing the [Y]* or *show the toy to Ben* to derive *[X] the [Y] to [NAME]*. Eventually, children may combine different schemes to form PD like sentence like *I'm [X]ing the [Y] to [NAME]* before arriving at fully abstract dative constructions. Some support for this comes from studies showing that children show a better understanding of prototypical datives that are consistent with such slot and frame patterns (Rowland & Noble, 2010).

2.2.5. Verb-structure links

In constructivists' views, children start off with constructions that are specific to individual verbs and they learn verb-structure links by observing the patterns with which verbs occur in different structures in the language input. Children are sensitive to frequencies with which different verbs occur in different structures and more frequently used verb-structure construction lead to stronger representation and thus more likelihood that those representations will be used in own language production. Campbell & Tomasello (2001), for example, demonstrated that children's use of verbs in different dative structures was highly similar to their parents'. They found that DO-dative uses of *give* outnumbered PO-dative uses of *give* by around 3:1 in both parental and child speech. Other verbs, such as *make*, showed a PO-bias in some parents and a DO-bias in others, with children generally replicating this pattern in their own speech, this way acquiring the biases with which different verbs are used in different structures.

A slightly different mechanism is described by Wonnacott, Newport, and Tanenhaus (2008). Their study showed that people could learn verb-structure links by tracking the occurrences and co-occurrences (e.g. transitional probabilities) of verbs and structures in the input. In a miniature artificial language study, the authors presented people with sentences where some of 12 verbs could occur in only one of two structures, while other verbs could alternate between both structures. The authors found that participants could track both verb-specific statistics, such as the probability of different verbs occurring in certain structures, and verb-general statistics, such as the likelihood of certain structure occurring across the verbs of the language. The two sources of information competed leading to complex patterns of behaviours in the way people made inferences about the use of different verbs in different structures based on their frequency and distribution of verb types across the language.

2.2.6. Summary: Constructivist approach to language acquisition

In sum, constructivist theories argue that learners exploit the regularities in the language input to extract and slowly build language structure. This process is supported by domain-general learning mechanisms that are sensitive to frequencies and transitional probabilities (Monaghan and Christiansen, 2010). These mechanisms help children abstract over different instances of language that share similar distributional patterns to form abstract representations that resemble syntactic categories, inflection rules, and other rule-like structural constraints.

3. Explaining linguistic adaptation with language acquisition theories

The main focus of language acquisition theories is to explain how children arrive at adult-like language knowledge at various levels. In other words, the goal is to explain how children progress from first basic utterances to complex error-free sentence constructions. This implies an end state of learning and the question is whether the same learning mechanisms could extend to linguistic adaptation phenomena that show that language continues to change in adults. More specifically, these theories should be able to explain verb bias and structural priming studies that show that continuous language input could change people's verb bias preferences or create long-term changes that bias the speakers to prefer certain sentence structures over others, based on their recent experience with those structures. The section below explores how such linguistic adaptation phenomena could be incorporated into the language acquisition theories described above.

3.1 Generativist theories and linguistic adaptation

Generativist theories like that of Pinker (1989) offer a comprehensive account of how verb-structure associations are formed through the learning of narrow-range rules. However, in this theory, once the child recognizes that a certain verb is a member of a certain semantic class, then learning is considered to be complete and there is no need to adjust these representations. Since the verb-structure behavior involves categorical class membership, the theory does not offer a mechanism to explain gradual changes in verb bias like those that are seen in linguistic adaptation studies.

Generativist theories also fall short at explaining structural priming effects, where the processing of one structure leads to changes that bias the speaker to repeat that structure in the future. Generativist theories are concerned with explaining how children learn adult-like language representations and once certain aspect of language is learned (e.g. word order parameter is set), no further learning is needed. Thus, generative theories assume that language learning takes place through categorical changes such as the assignment of a word to a category or the assignment of a parameter, and hence they are not well equipped for explaining the sentence-grain changes that are evident in linguistic adaptation studies.

3.2 Constructivist theories and linguistic adaptation

Constructivist theories of language acquisition are appealing in that their input-based mechanisms provide a close link between language input and language representations, which is critical for linguistic adaptation. However, a downside of these theories is that different studies focus on specific levels of language structure learning, which makes it difficult to incorporate the proposals into a single mechanism. For example, studies concerned with category learning theorise that people learn word-to-word transitions to form abstract categories. Phrase learning studies postulate that people learn category-to-category transitions. Verb subcategorization theories hold the view that people learn associations between verbs and structures. It is not clear how all these sources interact and how people make decisions about what type of information they must pay attention to. For example, in a verb bias study, Wonnacott et al. (2008) concluded that participants learned verb-specific regularities, which could be supported by transitions between words. However, they also learned verb-general regularities that could generalize to novel verbs and that were not based on transitional regularities between words. It seems that two different learning mechanisms support these two tasks and it is difficult to see how they would interact or complement each other. Also, in that specific study, the authors used a language that contained no determiners. Learners of English, for example, must learn that in the sentence *the boys give a book to the girl,* the verb *give* is followed by a noun phrase. It is not clear how the learner would decide whether to learn verb-structure relationships using transitions between phrase, category or word. This is further complicated by other sources of statistical learning information like calculation of backward transitional probabilities (e.g. Saffran, 2001), frame frequency computations (e.g. Mintz, 2003), schematization and abstraction (Tomasello, 2000).

Also, it is not obvious how constructivist theories could explain abstract structural priming and its lasting effects in adults. Slot-and-frame approach proponents, who emphasize the processes of abstraction and schematization, agree that encounters of perceptually similar language structures 'permanently change the user's linguistic representations in some way' (Abbot-Smith & Tomasello, 2006, p. 280). However, as the authors themselves agree, 'it is abundantly clear that we know very little about how all of this actually works in practice'. Indeed, these theories say little about how the mechanisms used by early language learners relate to language processing in adult life. Likewise, while statistical learning mechanisms seem consistent with structure adaptation in adults, it is still not clear how these mechanisms would explain the effects of priming that takes place at the level of abstract sentence structure. For example, how would the learning of transitional probabilities between the words in a PD sentence like "a man baked a cake for a princess" explain the biasing towards a PD structure, such that speakers are more likely to use that structure with a different set of words (e.g., *The boy gives a book to the girls).* Thus constructivist theories have gradual mechanisms that could support sentence-grain and year-grain learning, but the nature of the mechanism that supports the changes at various levels is still not clear.

3.3 Summary: Language acquisition theories and linguistic adaptation

To sum up, neither generativist nor constructivist language acquisition theories offer a clear mechanism that could extend to adult linguistic adaptation studies. The statistical learning approach is promising in that it captures the relationship between language input and changes in language representations. But statistical learning depends on the units that statistics are collected over. For example, there is evidence that infants collect statistics over syllables (Saffran, Johnson, Aslin & Newport, 1999), which is thought to support their ability to segment their input. But if these statistics are used in selecting words in on-line production, they will reduce the ability of speakers to create novel generalizations, because the syllable statistics in the novel generalization may not be attested in the input. Thus, statistical knowledge will depend on the level of representations that they are applied to and these regularities need to be balanced with mechanisms that support generalization. In the following section, we examine these issues within an explicit connectionist model, which is designed to learn statistical regularities at various levels but also can use meaning to generalize beyond the input. We will argue that this framework provides a way to apply statistical learning to both language acquisition and linguistic adaptation.

4. Linguistic Adaptation Mechanism of Language Learning (LAMOLL)

Connectionist models are useful for thinking about language acquisition and linguistic adaptation processes because they can simulate the learning processes that take many years in human participants. These models have been primarily developed to refute the claims that language learning requires language-specific learning mechanisms by showing that many language constraints could be induced directly from the language input. They instantiate the idea that statistical learning is at the heart of language learning but, in contrast to behavioural statistical learning studies, they provide an explicit mechanism that can be used to link the learning in adults and the learning in children.

4.1 The Dual-Path model as a model of language acquisition and linguistic adaptation

A model that assumes a close connection between language acquisition and linguistic adaptation is the Dual-path model (Dell & Chang, 2014). Developed as a connectionist model of language acquisition and adult language production (Chang, 2002), the model has been used to show that the same mechanism that helps it learn a language also gives rise to linguistic adaptation effects such as structural priming as a by-product of learning (Chang, Dell & Bock, 2006). The model uses prediction as the basis for learning. When it "hears" an incoming sentence, it attempts to predict the words in it on a word-by-word basis. When the predicted word mismatches the input word, an error signal is generated, which is used to change the weights in the model to improve the prediction abilities in the future. Importantly, the model is not told what aspects of the language it must learn and the representations that the model forms as the result of its prediction-based learning are emergent in nature. In other words, the model does not know about syntactic categories or structures. Instead, it learns these representations if they help the model predict the words more accurately. For example, in the language input word “the” could be followed by many other words like “boy”, “girl”, “book” and it is not always possible to predict which word exactly will be the correct grammatical continuation of the sentence. A compact way to make this prediction is to predict a category of NOUN, which is linked to all of the possible words. The model does this by gradually merging the representations for these different nouns and linking them to the previous word “the”. The model might also learn that nouns occur after the word “a” and that might cause it to develop a category of DETerminers which tends to predict the members of the NOUN category. Gradually over time, these categories could be combined into larger structures like PD and DO. Thus, error-based learning can explain how categories and structures are learned from a large mass of inputs.

Importantly, the same error-based learning mechanism was also able to explain abstract structural priming effects (Chang, Dell & Bock, 2006). In this study, the “child” model was trained on a large set of sentences until it learned dative representations. When the model reached adult-like performance, the “adult” model was given a PD prime sentence with the learning mechanism ON. This caused the model to adjust its representations to increase its expectation of the same structure again. As the result, when the model produced the target, it was more likely to use the primed PD structure. Such behaviour resembles abstract structural priming effects seen in humans. Indeed the model was able to explain a wide range of adult structural priming-related phenomena (Chang, et al. 2006; Chang, Baumann, Pappert, & Fitz, 2015) including structural priming effects that last over many structurally unrelated sentences, as in the human studies (Bock and Griffin 2000; Bock, Dell, Chang & Onishi. 2007).

The model has also been used to explain verb bias effects. Twomey, Chang, and Ambridge (2015), used the Dual-path model to explain how children learn to use verbs in locative structures. For instance, some verbs (e.g. *spray*) could occur in two different locative structures like *the woman sprayed water onto the wall* and *the woman sprayed the wall with water*. However, some verbs (e.g. fill) are restricted to only one of the structures like *The woman filled the salt shaker with salt* but not *The woman filled some salt into the shaker.* The model learned from distributional regularities that certain verb classes were restricted in their use and didn't automatically generalize their use to both structures. However, in contrast to Pinker's (1989) narrow range rules that defined semantic verb class as concrete categories, the classes in the connectionist model were also sensitive to the frequencies with which verbs occur in different structures. In other words, it learned that some verb classes were biased 75% towards one structure, while other classes were biased 25% to the other structure showing verb bias effects. Although the model was not used to explicitly test linguistic adaptation showing how verb biases change in adults, the model's representations would naturally respond to the changes in the language input, owing to the adaptive nature of its learning mechanism that is sensitive to frequencies.

In sum, the Dual Path model provides an explicit mechanism that could support both language acquisition and linguistic adaptation by treating the two as inextricably linked. Its assumption is that on-line processing of an individual instance of language input (e.g. sentence) creates small sentence-grain changes in its representations that result in processing biases like verb bias or structural priming as by-products of learning. Such small changes accumulate and result in year-grain long-term changes that become the structural representations of the language. Thus the same process supports both the formation of language representations from scratch and subtler changes that manifest as linguistic adaptation effects in adults. Its prediction-based learning mechanism naturally helps the model develop representations that reflect the distributional properties of the language and produce behaviours that correlate with the statistical properties of the language described by the statistical learning studies. In light of this, the model provides a compelling account of language learning as the linguistic adaptation that could help bridge the learning effects seen in children and adults. These assumptions will be used as the basis for LAMOLL account in this thesis.

Before we proceed with the testing of the LAMOLL account, it is important to distinguish it from other prediction-based theories of linguistic adaptation phenomena like structural priming. One such account is Surprisal Sensitive-Persistence hypothesis by Jaeger and Snider (2007). The authors hold the view that language-processing system is set up in such way that it implicitly maintains and updates probabilistic distributions over linguistic structures. Such maintenance of probability distributions in an inherent part of language processing system and the persistence of structural priming, under this account, is seen as a correlate of the maintenance of syntactic probability distributions. For example, whenever it encounters an instance of NPNP structure, it takes it as a piece of evidence that affects the structure's probability distribution (e.g. the distribution of NPNP vs. NPPP structure). Less probable syntactic structures lead to a bigger change in the probability distribution, which in turn leads to an increased probability of reusing the same structure. In theory, this account leads to similar predictions when it comes to explaining structural priming effects. However, the Surprisal Sensitive-Persistence hypothesis offers no explanation of how language structures are learned in the first place, which is a critical feature of any model that aims to link the two processes.

4.2 Motivation for the present studies

While the Dual-Path model can explain a range of linguistic behaviors in humans, it is a connectionist model and its assumptions are yet to be validated behaviourally in humans. Its psychological reality critically depends on several assumptions. In the model, sentence-level effects like structural priming, continuous adaptation to language input, and long-term language structure acquisition effects are supported by a common learning mechanism. In other words, all these seemingly different phenomena result from prediction-based learning associated with the processing of individual sentences.

In the existing studies, there is evidence for adaptation/learning at various grains in these studies. For structural priming studies, both in natural languages (e.g. Bock, 1986) and in artificial languages (e.g. Fehér et al, 2016), there is evidence that changes that occur during the processing of the prime sentence can impact the processing of the following sentences, which we call sentence-grain changes. There are also studies that show that participants acquire an artificial language (Wonacott et al. 2008) or particular experimental constraints from the input over a whole experiment in natural language (Dell, Reed, Adams, & Meyer, 2000; Warker & Dell, 2006) and we will call this experiment-grain learning. Finally, language learning takes place over multiple years in the real world and we will call this year-grain learning. However, while there is evidence for adaptation/learning at each of these grain sizes, linking these processes as supported by a common learning mechanism is a complicated task. For instance, how sentence-grain changes are related to experiment-grain changes are complicated by the studies that treat language learning and linguistic adaptation as separate processes. There are studies that show that participants exhibit structural priming-like effects after exposing them to simple artificial languages (Fehér et al., 2016). However, they use blocked designs where learning and testing are treated as different tasks. Fehér et al. (2016), for instance, used a task where participants were trained on a novel language by presenting pictures of novel objects together with text descriptions to teach them the meaning and word order of the language. After the training, structural priming was tested in a separate task where participants interacted with a computer in a game that required identifying the correct scene based on the computer's description or describe a scene for a computer to do the same. Priming was assessed by looking at the proportion of the description where participants aligned the structure of their sentence with the computer's description. Such a method does not rule out the possibility that language learning and structural priming are unrelated processes supported by different mechanisms.

Another issue is to link experiment-grain changes with year-grain learning, as is predicted by the LAMOLL account. Due to practical issues in doing a multi-year study, it is difficult to create a controlled experiment that would assess learning over many years. The Dual-path model has been able to explain the sentence-grain adaptation of structures and it can explain some aspects of year-grain learning of language but it has not been used to explain the detailed changes that occur over years in different populations. One of the critical assumptions of LAMOLL account is that linguistic adaptation in adults relies on a mechanism that supports language acquisition in children. However, second language (L2) learning studies show that as language learners get older, the language learning ability seems to decline due to a sensitive period that is characterized as the time window until about puberty where language learning is most effective (Curtiss, 1977; Johnson and Newport, 1989; Lenneberg, 1967; Long, 1990; Newport, 1990). Some studies have demonstrated that the knowledge of the second language (L2) structure in older L2 learners does not correlate with length of language exposure as measured by years spent in the L2 environment (e.g., Johnson & Newport, 1989). Since linguistic adaptation and learning critically depend on extracting linguistic regularities from the language input, these studies challenge the notion that learning processes in children and adults could be linked as supported by a common mechanism.

The goal of the present work is to test the LAMOLL account by addressing the limitations above. Namely, the following series of studies will aim to demonstrate that a) the mechanism described by the connectionist model is psychologically real and supports sentence-grain and experiment-grain learning processes in human participants; b) the same mechanism that explained adaptation and language acquisition in Chang el at. (2006) could also explain the development changes in L2 learners who started learning a language at different ages and who received different amounts of language exposure.

5. Structure of the thesis

The current thesis reports three separate studies that constitute stand-alone pieces of work that are either published or are in preparation/submitted and that are accompanied by their own introduction motivating the experimental part of the studies and discussion exploring the implications of the results. Each study has implications for the LAMOLL account and their findings will be summarized and discussed in the context of the account in the discussion of the thesis. The rest of the thesis will follow the following structure:

Chapter 2 focuses on the prediction of the Dual-Path model that on-line processing of language input is linked to language structure acquisition and linguistic adaptation effects. Using a non-linguistic artificial grammar learning (AGL) task, a series of four studies aim to investigate the learning of PD and DO-like structures and linguistic adaptation effects associated with the processing of those structures, like verb bias and structural priming, using detailed on-line measures of reaction times and accuracy during learning. In line with the predictions of the model, the study also attempts to demonstrate that learning in these tasks is supported by a prediction based learning mechanism. This study will be used to provide the link between sentence-grain size and experiment-grain size learning effects.

Chapter 3 focuses on the model's prediction that language learning effects in adults and children could be linked to a common learning mechanism. It will investigate the sensitive period effects in L2 to address the lack of correlations between the length of language exposure and adult L2 grammar knowledge. The study will present a version of the Dual Path model to show that the same mechanism that is used to learn the first language and that can explain structural priming effects can explain sensitive period effects in adults. The study will be used to provide support for the long-term effects of the LAMOLL mechanism and the commonalities between the learning processes in adults and children.

Chapter 4 will use the AGL method developed in Chapter 2 to further explore the results reported by the connectionist model in Chapter 3. It will test the extent to which controlled AGL studies could be used to understand language learning effects in natural language, given the commonalities in their learning mechanisms. The studies will explore the negative effect of frequency reported in Chapter 3, where performance with grammar rules like determiner, plural, past tense and third person singular in adults was found to be negatively associated with the frequency with which they occur in language input. Such findings are at odds with input-based theories and the study will attempt to replicate and understand these effects at the sentence-grain and experiment-grain level. The studies will also look at structural priming effects associated with rule learning.

Finally, Chapter 5 will provide a general discussion explaining how the results from all three studies support the LAMOLL account and will offer some possible directions for the future research.

Chapter 2: Linking acquisition and adaptation: A non-linguistic study of structure acquisition, verb bias, and structural priming

The following series of studies are currently revised for publication in collaboration with Andrew Jessop and Franklin Chang.

1. Introduction

It is known that linguistic representations in adults change in response to experience and such *linguistic adaptation* effects occur at various levels of linguistic structure (Dell, Reed, Adams, & Meyer, 2000; Oppenheim, Dell, & Schwartz, 2010; Bock, Dell, Chang, & Onishi, 2007). It is also the case that children must learn these linguistic representations during language acquisition. However, while both phenomena involve changes in language representations in response to the linguistic input, it has been traditionally thought that very different mechanisms supported these processes. For example, Chomsky (1981) postulated that children learn a language by setting a combination of mental syntactic parameters that determine the way that syntactic structures are generated (e.g., verb direction parameter determines the placement of the verb). Once the parameters are set, language representations no longer change. In contrast, theories of linguistic adaptation often provide mechanisms that explain how syntactic structures change, but do not specify how the structures are learned in the first place (e.g., Jaeger & Snider, 2013; Perfors, Tenenbaum & Wonnacott, 2010). The aim of the present paper is to explore whether similar mechanisms can explain both language acquisition and adult linguistic adaptation phenomena.

One prominent linguistic adaptation effect is structural priming (Bock, 1986). Structural priming studies involve structural alternations, where participants could choose between different sentence structures to express their message. For example, the dative alternation involves a choice between the double object (e.g. *the man gave the woman the book*) and prepositional dative (e.g. *the man gave a book to the woman*). In structural priming tasks, the structure that is chosen can be influenced by an earlier heard or produced prime sentence. For instance, if participants hear a prepositional dative structure sentence like '*Mary sent the package to her mother*', they are more likely to use the prepositional dative (e.g., *the man gave a book to the woman*) in their own utterance. Importantly, these priming effects persist over time (Bock et al., 2007; Bock & Griffin, 2000) and thus it has been argued that some type of learning mechanisms supports these adaptive changes in adult speakers (Chang, Dell & Bock, 2006). Consistent with these accounts, Jaeger and Snider (2013) showed that structural priming effects could be explained by a rational expectation learning mechanism (see also Myslín & Levy, 2016) that is sensitive to the statistics in the environment. But to calculate the initial expectations in the environment, they used human-labelled corpora where utterances were labelled with syntactic structure labels. Therefore, this theory uses different mechanisms to explain linguistic adaptation (rational expectation-based learning) and the identification of the syntactic structures (human labelling of utterances). The use of different mechanisms for explaining how knowledge is learned and how it changes in adults is wide-spread in the literature (e.g., Fine & Jaeger. 2013; Reitter, Keller & Moore, 2011). The main issue in this work is to examine whether a common mechanism could support both adaptation of existing structures and the acquisition of structures in the first place.

While linguistic adaptation studies often overlook the aspects of language acquisition, theories of language acquisition do not easily extend to explaining the learning effects associated with linguistic adaptation. For example, some language acquisition theories assume that language acquisition is supported by the inborn knowledge of linguistic categories, structures, rules, parameters, and principles (Chomsky, 1975; Pinker, 1984; Valian, 1986). The goal of learning is to map the language input onto the existing representations to activate the right set of rules that are required to process that specific language. One example of this approach is Yang's (2002) variational learning model. Under his account, there are various binary mental parameters that control different aspects of structure generation (e.g. verb second V2 parameter determines the position of the verb in a sentence). During the language acquisition, these parameters generate a range of different grammars and these grammars compete for the best match to the language input before the right combination of parameters is set. However, while this mechanism is sensitive to distributional regularities in the input, the goal of the mechanism is to assign the right parameter values to ensure that the target language is learned. It is not clear how this model could explain linguistic adaptation in adults. If the parameters were left unset in adults, then competing grammar could lead to errors if, for example, a language user found himself/herself in an environment where they did not hear a passive structure, which would suggest to the system that the structure is not grammatical. In addition to these issues, this approach also uses language statistics from the corpora that have been labelled by linguists, so it does not provide a theory of how the child links utterances to the existing grammar knowledge in the first place. It suffers from the problem of circularity raised by Mazuka (2014) that such generative acquisition theories often require a parser to get the categories which are used to set up a language-specific parser.

In contrast to the nativist accounts that rely on pre-existing linguistic representations, a range of theories argue that language acquisition can be supported by domain-general learning mechanisms like statistical learning (Saffran & Thiessen, 2007). According to these theories, learners collect statistics on linguistic units like syllables or words and use these statistics to discover higher order representations like words or categories (Hunt & Aslin, 2010; Mintz, 2002; Saffran, Newport & Aslin, 1996a; Saffran, 2001). However, the challenge for these approaches is to explain the development of abstract syntactic structures. A key feature of abstract structures is that they can be applied to elements that have never appeared in these structures before. For example, if a child learns the word '*Pokemon*', she can use it in a sentence like '*The boy gave to his dad the Pokemon that he caught yesterday*', even if she has never heard this word in a heavy NP shifted dative structure with a relative clause. While there is a lot of evidence for statistical learning in language acquisition, it has been difficult to trace all of the steps that yield abstract syntactic structures and in turn explain the adaptation of those structures in adults.

One approach that could overcome these limitations is provided by a connectionist model called the Dual Path model (Chang, Dell, & Bock, 2006). In their study, the model started out with no knowledge of English syntax. To teach the language, the model was given English sentences along with the corresponding message (the meaning). The model had to learn English syntactic representations that mapped between the meaning of the sentence and the correct sentence structure. To test the model on structural priming, it was presented with a sequence of words of the prime sentence (no structure or message information was present) with the learning left ON. Then it was given a different message and it had to generate a sentence of its choice. The authors found that the processing of the prime sentence cause the model to adjust its linguistic representations in the way that made the model bias towards that structure. This made the model more likely to use that structure in its own sentence production. Since the prime and target had different words and meaning, the fact that the model showed structural priming is evidence that it had learned abstract syntactic structures during the language acquisition. Thus in contrast to the existing accounts of linguistic adaptation and language acquisition, this model demonstrated that the same computational mechanism that was used to learn syntactic structures like DO and PD could also explain linguistic adaptation in adults.

Although Chang et al. (2006) model provides a computational bridge between acquisition and linguistic adaptation, there is still no behavioural evidence that clearly demonstrates this link in humans. The study that comes closest is Fehér, Wonnacott, & Smith (2016), who taught participants a novel artificial language that consisted of four simple possible sentence structures: Numeral-Noun, Noun-Numeral, Adjective-Noun, and Noun-Adjective. In their study, participants first learned the vocabulary of the language by learning the associations between novel words and pictures of objects referring to them. They were then taught different sentence structures by presenting images depicting the objects in different numbers (numeral information) or different colour (adjective information) along with 2-word descriptions of the image in one of the possible orders. The training task was followed by a test phase where participants interacted with a computer partner taking turns in describing objects presented by a partner and then selecting objects based on partner's description. The authors found that participants were sensitive to the choice of structure of their partner where they tended to reuse their structure in their own sentences. This showed that participants were able to learn a novel structural alternation during the training phase and then exhibited structural priming effects in the test phase. However, the blocked design of the study where tasks are separated into language learning and language processing phases implies the view that acquisition and adaptation are different tasks. This does not allow linking the mechanism of structural priming to learning because structural priming may be a separate phenomenon that manifests when people learn a language. For instance, it may be the case that structural priming is based on residual activation (e.g. Pickering & Branigan, 1998) of the recently used structure that is based on a mechanism that is not related to learning. Also, the social nature of the task means that structural priming could be the result of the interactive alignment, which according to Pickering & Garrod (2004) simplifies production and comprehension in dialogue and is based on an interactive inference mechanism that enables development of local dialogue routines that simplify language processing. Lastly, people may have aligned their structure using an explicit strategy assuming that the computer partner knew the language better and felt the pressure to use similar word order in their description. Thus the use of a blocked design where acquisition and priming were in separate blocks in Fehér, Wonnacott, & Smith (2016) does not allow drawing inferences about the relationship between structural priming in language acquisition.

If the same learning mechanism is involved in both phenomena, then it should be possible to develop an on-line task that allows examining both learning and priming simultaneously, as in artificial grammar serial reaction time studies. For example, Cleeremans and McClelland (1991) exposed participants to sequences of dots appearing in six locations on a computer screen and participants had to produce the sequences by clicking on a corresponding key for each screen position. Unbeknown to the participants, the sequences followed the rules in a finite-state grammar and the authors tested if people became sensitive to the probabilistic relationships between the elements of the language. Unlike the blocked designs, the task allowed measuring detailed reaction time and accuracy responses, as they were learning/processing the language. The authors found that participants performed faster when processing more predictable transitions as compared to the novel and ungrammatical transitions that violated the grammar. They also found that participants' performance improved over the course of the study, showing that they were adapting to the incoming information and adjusting their representations of the language structure. Importantly, the training items also acted as test items which made it possible to take on-line measure of learning and processing at the same time. However, while there are many studies using serial reaction time paradigms to examine artificial grammar acquisition (for reviews, see Clegg, DiGirolamo, and Keele, 1998; Abrahamse, Jiménez, Verwey and Clegg, 2010), these studies have typically not used structures that closely resemble those in human languages and it can be difficult to generalize from these studies to natural language. In the present work, we use an on-line serial reaction time task to examine the acquisition of an artificial language that is modelled on the English dative alternation.

In summary, we examine whether similar mechanisms can support language acquisition and linguistic adaptation within a serial reaction time study. Specifically, we are interested whether participants can go from having no knowledge about the language to encoding abstract structures that mimic the English dative alternation. During the on-line learning of this language, we test linguistic adaptation phenomena such as structural priming to see if participants implicitly change their biases for these structures under the same conditions as they use for learning the structures. Since we are looking at the learning of these structures on-line, we will be collecting multiple measures such as reaction times and accuracy to understand the nature of the developing representations. In the next section, we will review the literature around several dative phenomena that we will examine in this task with a particular emphasis on how they influence different dependent measures.

1.1. Acquisition and adaptation of the Dative Alternation.

The dative alternation involves the double object (DO) and prepositional dative (PD) structures that convey approximately the same meaning. The double object dative structure typically has a dative verb followed by two noun phrases (NP VERB NP NP, e.g., *the woman sent the man the book*), while the prepositional dative will have one noun phrase and a prepositional phrase in post-verbal position (NP VERB NP PP, e.g., *the woman sent the book to the man*). There is not a lot of work that directly tracks how the dative structures themselves are learned. It is difficult to do this because dative structures build on transitive structures (e.g., *the woman sent the book*) and it can be hard to identify a unique point when the dative structure becomes distinct from the transitive. Snyder and Stromwold (1997) use the first clear usage of each structure as evidence for the acquisition of that structure. But this means that if a child memorizes an idiom or a lexical frame with a slot (e.g., *give me a X*), this criterion would imply that the structure has been acquired, even if the child cannot generalize it to any other verb. Campbell and Tomasello (2001) found that before 3-years of age, children tended to use the dative structures with a limited set of verbs initially and most dative verbs were initially used in transitive structures. Gropen, Pinker, Hollander, Goldberg & Wilson (1989) found that 95% of double object dative tokens were likely learned from adult input suggesting that early knowledge of the dative could be quite conservative. But Conwell and Demuth (2007) found that three-year-old children could alternate with a novel verb at 3 years of age when given some priming for both structures.

There is relatively more work on the relationship between verbs and dative structures. Young children often overgeneralize verbs to dative structures that are not allowed in the adult grammar (e.g., *I will brush him his hair*; Mazurkewich & White, 1984). Pinker (1989) explains this in terms of broad range rules that allow verbs that are heard in prepositional datives (e.g., *I will brush his hair for him*) to be converted into verbs that can appear in double objects. Gradually, children learn the narrow range rules which are learned semantic constraints on particular verbs that govern their acceptability in particular structures. Gropen et al. (1989) tested this theory and found support for the view that children and adults were sensitive to these constraints on verb-structure links. This work suggests that learning verb-structure links involves specialized knowledge (broad/narrow range rules) which are specific for the acquisition of the verbs in the dative alternation.

Verb-structure links are not simply binary but instead are often graded in nature. In production norming tasks (Garnsey et al., 1997; Trueswell, Tanenhaus, & Kello, 1993), participants are asked to complete sentence fragments and they produce structures that occur more frequently with the verb. The graded nature of these biases is thought to arise from the frequency distribution in corpora (Lapata, Keller, & Walde, 2001). These effects also occur in on-line production tasks like Lombardi and Potter (1992) who found that recall of a target sentence (e.g., "the author gave the library a letter") could be biased to the alternative structure by hearing a verb that was biased towards that structure (e.g., "donate"). Verb bias also increases production speed (Jennings, Randall, & Tyler, 1997; Gahl & Garnsey, 2004; Stallings, MacDonald, & O'Seaghdha, 1998). Verb bias effects are also evident in comprehension studies where processing is faster when verbs occur in preferred structures in self-paced moving window reading (e.g., Kennison, 2009, exp. 1) and eye tracking (e.g., Garnsey et al., 1997, exp. 1). Likewise, these effects also manifest in comprehension accuracy, where matching verb-structure pairs yield fewer errors (Gahl, 2002) or higher accuracy in comprehension questions (Gennari & MacDonald; 2009). The graded nature of verb bias is consistent with frequency sensitive learning mechanisms like statistical learning. Wonnacott, Newport, and Tanenhaus (2008) found that verb-structural regularities in an artificial language influenced production, judgements, and on-line processing. Twomey, Chang, and Ambridge (2014) found that the distribution of noun phrases in post-verbal position was a useful cue for identifying the structural preferences of locative verbs and they showed that learners could use these cues to learn verb bias in an on-line task.

Another phenomenon which is often tested with the dative alternation is structural priming, which is a tendency to reuse previously heard syntactic structures. Structural priming occurs in various production tasks like picture description (e.g. Bock, 1986) or sentence completion (e.g. Pickering & Branigan, 1998). It also occurs in sentence recall tasks, where the prime structure changes the structure that the participant is trying to recall (e.g., Potter & Lombardi, 1998; see also Chang, Bock & Goldberg, 2003; Tooley & Bock, 2014). Priming can also increase the speed at which sentences are produced (e.g. Segaert, Menenti, Weber, & Hagoort, 2011; Smith & Wheeldon, 2001; Wheeldon & Smith, 2003). It is also evident in reaction times in comprehension tasks using self-paced reading (Weber & Indefrey 2009), eye-tracking reading (Traxler, 2008), or visual world eye-tracking (Arai, Van Gompel, & Scheepers, 2007; Thothathiri & Snedeker, 2008). Tooley & Bock (2014) directly compared priming in self-paced reading and RSVP production and found that the magnitude of priming was similar across both modalities. It is less clear if priming increases comprehension accuracy as many studies have not investigated this and some studies report no effect (e.g. Ledoux, Traxler, & Swaab, 2007).

In summary, different theories and methods have been used to study the acquisition of the dative alternation, acquisition and use of verb-structure links, and structural priming. In order to see if these different phenomena can be integrated together under a common set of mechanisms, it is necessary to have a task where changes in structures and verb-structure links can be monitored during acquisition and processing. Critically, as these phenomena have different properties in different dependent measures in comprehension and production, it is necessary to develop ways to distinguish how linguistic knowledge is used in different tasks.

2. Experiment 1: Verb bias and priming in the Circle Task

Many of the existing artificial grammar learning studies have involved presenting participants with recordings of strings of nonsense words (e.g. glim, blergen, tombat, ka) in certain probabilistic combinations dictated by grammar rules of an artificial language (e.g. Wonnacott et al., 2008). Typically, at the end of the training participants receive a forced choice grammar judgment task, where they have to distinguish strings that are permissible in the language from those that are not. Their ability to recognize grammatical strings is taken as evidence that they have learned the distributional regularities of the language (e.g. Aslin & Newport, 2012). However, the use of auditory word-like strings in these studies means that the activation of language-specific learning mechanisms cannot be ruled out in these tasks. For example, Marcus, Fernandes, & Johnson (2007) showed that people learn rules that are presented in speech format faster than when they are presented using non-speech materials (musical tones, animal sounds or varying timbres).

Furthermore, participants have to learn the words before they can begin to learn syntactic knowledge, and this adds noise and complexity to the study of syntactic knowledge within these tasks. To reduce these problems, we will use a serial reaction time (SRT) task, where participants learn to produce motor sequences based on visual cues (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987).

Our SRT task was inspired by Hunt & Aslin's (2010) study, where participants were presented with a response frame on a touch screen that contained 16 shapes forming a square. Next to the response frame was a stimulus box, where shapes were presented in various sequences one at a time. The task was to find each shape presented in the stimulus box in the response frame by tapping it as fast as possible. The different shapes formed distinct categories that occurred in various permitted combinations that conformed to a grammar. The transitions between the elements of different categories were probabilistic and could not be predicted with absolute certainty. Some transitions were withheld during the training to examine if people could generalize their knowledge at the test. They found that participants were faster at finding the shapes when the sequences followed the rules of the grammar as opposed to a random order. This increase in performance suggests that participants were able to extract distributional information from the input by tracking the probability that certain elements appear in adjacent positions (Hunt & Aslin, 2001). Furthermore, Hunt and Aslin (2010) have demonstrated that participants go beyond distributional information and respond faster to novel transitions that conform to the abstract grammar. They argued that this demonstrated abstract category knowledge.

Since our goal was to examine syntactic acquisition and adaptation, we simplified Hunt and Aslin's method in several ways. Instead of using arbitrary shapes that had to be learned before sequence learning could take place, we used letter symbols that acted as ‘words’ in the artificial language. Since participants were familiar with letter symbols, they could quickly begin to learn the sequencing regularities. Also, for practical reasons we used a computer mouse-based task, where participants selected each letter by moving a mouse cursor on top of it. To equate the distance from the mouse cursor to the letters, the cursor appeared at the centre of the screen at the beginning of each trial and the letters were arranged in a circle around it. Thus we call this the *Circle Task*. Stimulus letter sequences were presented in the centre one letter at a time. The participants had to move the mouse to the matching letter on the circle, while their reaction times and accuracy were recorded. We expected participants to become faster and more accurate as they acquire the grammar (Cleeremans & McClelland, 1991; Hunt & Aslin, 2001, 2010), which mirrors the increase in fluency that occurs over language development (e.g. Ellis & Wells, 1980; Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002).

The main difference between this and the previous studies is that the language in the present study was based on English sentence structures. To simplify learning, the grammar contained no articles. For example, the English prepositional dative (PD) sentence ‘*The man sent the dress to the woman*’ could be rendered as a letter sequence X J F C H, where X was *man*, J was *sent*, F was *dress*, C was *to*, and H was *woman*, while the double object (DO) dative sentence “The man sent the woman the dress” could be expressed as X J H F. We created verb bias by having two dative verb categories (DVERBD and DVERBP). DVERBD category occurred 75% in DO structure, while the DVERBP category appeared 75% in PD structure. Since DO and PD structures differ in terms of which category follows the verb (verbs in DO structures are followed by animate nouns, while verbs in PD structures are followed by inanimate noun category), we could test verb bias effects by looking at the participants' ability to find the post-verbal letter. If participants learn and become sensitive to verb bias, then they would be expected to be faster at finding the postverbal letter when the structure of the sequence matched the verb's bias, as in comprehension studies in natural language (Garnsey et al., 1997; Jennings et al., 1997; Kennison, 2009). Based on the findings that sentences are comprehended more accurately when sentence structure is consistent with verb bias (Gahl, 2002; Gennari & MacDonald, 2009), we predicted that people would also make fewer mistakes when selecting the post-verbal letter when sequence structure was consistent with the verb's bias.

In order to test structural priming, PD and DO structure sequences were presented in pairs in all combinations, where one acted as a prime and the other was a target. Based on natural language studies (e.g. Arai et al., 2007; Thothathiri & Snedeker, 2008), we hypothesized that people would be faster at finding the post-verbal letter in target sequences that are preceded by the same structure sequence than when the structures are different. Importantly, the prime and target sequences were made up of different letter sequences, so the priming benefits could not be explained by transitions between specific letters but instead provide evidence that participants have learned abstract structures.

2.1. Method

2.1.1. Participants

Data were collected from 78 participants, all of whom were recruited from the undergraduate student population at the University of Liverpool and received course credit for their time.

2.1.2. Materials

Letter sequences were generated from a grammar that was modelled on English sentences. The language was created using 7 word categories (Table 2.1): animate nouns (ANOUN), inanimate nouns (INOUN), intransitive verbs (IVERB), transitive verbs (TVERB), dative verbs with DO structure bias (DVERBD), dative verbs with PD structure bias (DVERBP), and preposition (PREP). Each category was associated with one to four letter symbols. The grammar combined these categories to create sequences that resembled English intransitive (IN), transitive (TR), double object dative (DO) and prepositional dative (PD) sentences (Table 2.2). Articles were removed from the language to make learning easier.

Table 2.1.

Category types, category, and symbols used to create the language

Category Type	Category	Symbols
Animate Noun	ANOUN	X, M, Y, H
Inanimate Noun	INOUN	F, Z, Q, P
Intransitive Verb	IVERB	W, L
Transitive Verb	TVERB	S, G
Dative verb with PD bias	DVERBP	J, B
Dative verb with DO bias	DVERBD	D, N
Preposition	PREP	C

Table 2.2.

Rules that were used to create letter strings of four different structures

Type	Category	English-equivalent example
IN	ANOUN IVERB	Boys sleep
TR	ANOUN TVERB INOUN	Boys like girls
DO	ANOUN DVERBD/DVERBP ANOUN INOUN	Boys gave/threw girls books
PD	ANOUN DVERBD/DVERBP INOUN PREP ANOUN	Boys gave/threw balls to girls

Participants were exposed to 120 sequences divided into 5 sections of 24 items. Each section had 8 items that tested structural priming in all combinations twice (PD-PD, PD-DO, DO-PD, DO-DO). In each pair, the first sequence acted as a prime and the second was the target in which priming effects were measured. Each prime-target pair was separated by a filler that had either IN or TR structure, which is common in structural priming studies (e.g. Bock, 1986).

Verb bias was manipulated by pairing DVERBD and DVERBP category letters with DO structure 75% of the time (6 times in each section), and PD structures 25% of the time (twice in each section). All target sequences contained the verbs that were biased towards that particular structure to reduce the effect of verb bias on structural priming in target sequences. Verb bias was manipulated in prime sequences (Table 2.3).

Once the language was created, the sequences were presented to all participants in the same order. Individual letters in the sequences were generated by randomly selecting the letters from appropriate categories and differed for each participant. No letters were repeated in adjacent sequences to make letter distribution more equal. This also ensured that structural priming effects could not be attributed to the repetition of the letters.

Table 2.3.

Example distribution of structures and verb bias in one section of the experiment

Structure	Match with verb's structural preference	Item type
IN	-	Filler
PD	Match	Prime
DO	Match	Target
TR	-	Filler
DO	Match	Prime
DO	Match	Target
TR	-	Filler
PD	Match	Prime
PD	Match	Target
IN	-	Filler
DO	Match	Prime
PD	Match	Target
IN	-	Filler
PD	Mismatch	Prime

DO	Match	Target
TR	-	Filler
DO	Mismatch	Prime
DO	Match	Target
TR	-	Filler
PD	Mismatch	Prime
PD	Match	Target
IN	-	Filler
DO	Mismatch	Prime
PD	Match	Target

2.1.3. Procedure

Participants were tested on PC desktop computers (15in screen, USB mouse, PS2 keyboard) running a Circle Task program created using Processing software (version 2.2.1, www.processing.org). Participants were tested in a quiet computer room, with up to six individuals on separate computers per session. Instructions were presented on a computer screen explaining that they were going to perform a letter-matching task and that they had to complete it as fast and as accurately as they could. They were instructed to use a mouse with one hand while operating the keyboard whenever required with the other hand.

At the start of the experiment, participants saw a blank screen saying ‘press SPACE to continue’. Once ready, pressing the space bar key started the experiment. Participants saw a circle of randomly distributed letters (Fig. 2.1) and the first letter of the stimulus string presented in the centre. All participants saw the same random distribution of letters on the circle. The mouse cursor started in the centre of the screen behind the stimulus letter. The task was to find the matching letter on the circle by moving a mouse cursor on top of it. Their response was indicated by highlighting the letter that was the closest to the mouse cursor when it reached the circle. This reset the cursor back to the centre where the next letter of the sequence was presented. Each letter sequence was separated by a blank screen saying ‘press SPACE to continue’. Participants were told that they could take a break when they saw the SPACE screen or they could continue by pressing the space bar. The experiment took approximately 15 minutes to complete.

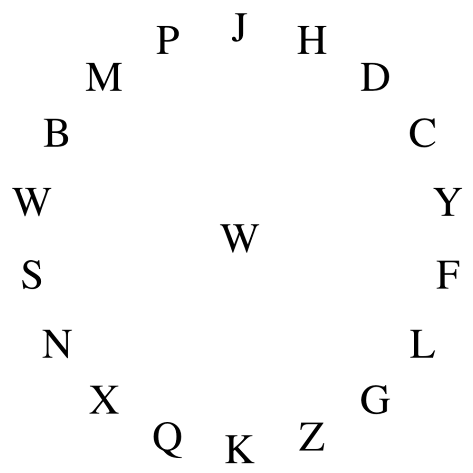


Figure 2.1. Visual display for Circle Task

2.1.4. Data and Statistical Analyses

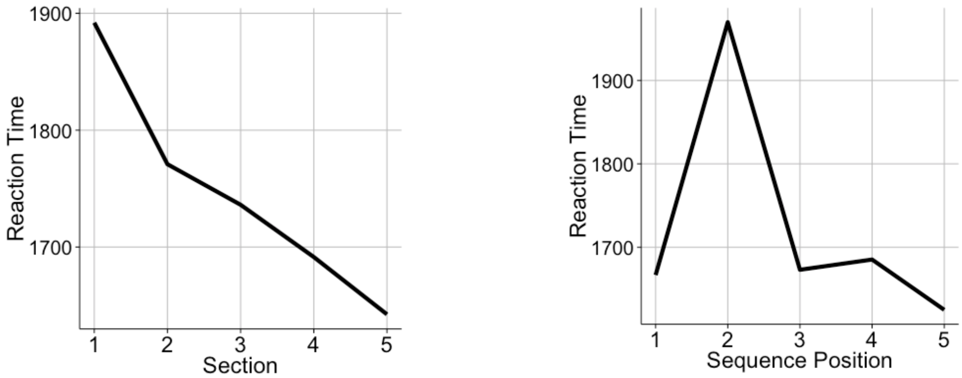
Reaction times in milliseconds were recorded for the time taken to move the mouse cursor from the centre to the correct symbol on the circle of letters. The error was defined as the selection of any letter that was not a target letter. The task produced a total of 36,339 responses with an error rate of 5.7%.

Only correct responses were used in reaction time (RT) analyses. RTs were log transformed and 2 standard deviations below and above the mean were removed as outliers to normalize the data. Reaction times were analysed using mixed effects linear regression. Accuracy analyses used error as the dependent measure and were analysed using mixed effects logistic regression. Both reaction time and accuracy analyses included participants as a random effect. Random effect structure (random slopes) was defined for each model separately (Barr, Levy, Scheepers, & Tily, 2013). P values were obtained through model comparison with likelihood-ratio test.

2.2. Results

In all of our studies, three separate analyses were performed to examine grammar acquisition, verb bias, and structural priming. To test if participants learned the grammar, we analysed how overall speed and accuracy changed at different sections and different sequence positions. If participants learned the grammar, they should become faster and more accurate in later sections of the experiment. Furthermore, if they learned some of the dependencies in the sequences, they should be faster to select items at later positions in sequences than those in earlier positions. Reaction times and accuracy were submitted to separate mixed models with section (centred) crossed with sequence position (centred) as predictor variables. The models included random slopes for the same two variables.

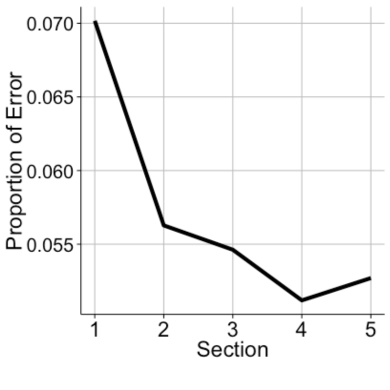
As seen in Fig. 2.2 (a), participants became increasingly faster at finding letters on the circle as the experiment progressed, β=-0.04, SE = 0.005, $\chi^2(1)$= 41.21, $p < .001$. Participants showed evidence of position-specific processing biases, as they were faster in later sequence positions (Fig 2.2, b), β=-0.02, SE = 0.002, $\chi^2(1)$ = 44.42, $p < .001$. No other effects were observed.



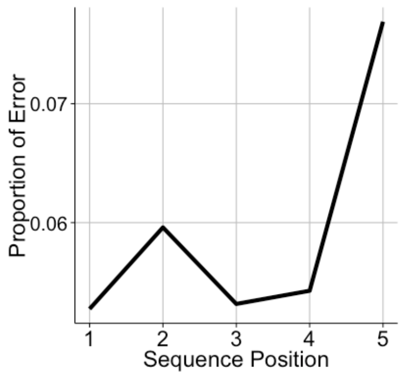
(a) **(b)**

Figure 2.2. Average reaction times in different sections (a) and at different sequence positions (b)

The *accuracy* analysis showed that participants became more accurate as the experiment progressed, β= -0.09, *SE* = 0.02, $\chi^2(1)$ = 12.1, $p < .001$. Despite the trend seen in Figure 3 (b), there was no main effect of sequence position (p = .1). Participants developed stronger position-specific expectations over the study, as shown by the growth of their errors in later sequence positions, β= 0.04, *SE* = 0.01, $\chi^2(1)$ = 6.72, $p = .01$.



(a)



(b)

Figure 2.3. Average proportion of error in different sections (a) and at different sequence position (b)

It is possible that speed improvements in this task were due to general task-based practice effects (e.g., faster mouse control) or task-specific learning (e.g., learning about the location of particular letters). To further demonstrate that participants acquired grammar-specific knowledge, we analysed RTs only for INOUN category letters, which occurred in TR, PD and DO structures. Since the task simply requires participants to recognize the letter presented in the centre and then move the mouse towards the letter on the circle, general or task-specific practice effects would affect participants' performance with INOUN letters in the same way, regardless of the structure they appeared in. However, if participants learned the structural information of the language, the predictability of INOUN category letters would differ for different structures and this difference should be reflected in reaction times. To test this, reaction times taken to find INOUN letters were predicted by section crossed with structure (as a factor). The model included random slopes for section and structure. Figure 2.4 shows a general learning effect ($\chi^2(1)$ = 40.91, p < .001) where reaction time improves over each subsequent section. In addition, there was also a main effect of structure, $\chi^2(1) = 15.01$ (2), $p < .001$, and an interaction between section and structure, $\chi^2(1) = 10.51$ (2), $p = .005$, indicating an increased differentiation between the structures as the participants learned the language. This shows that the task induced learning of the grammar beyond simple learning of INOUN letter locations on the circle. Accuracy analyses were not conducted, because there were few errors on INOUN letters in different structures.

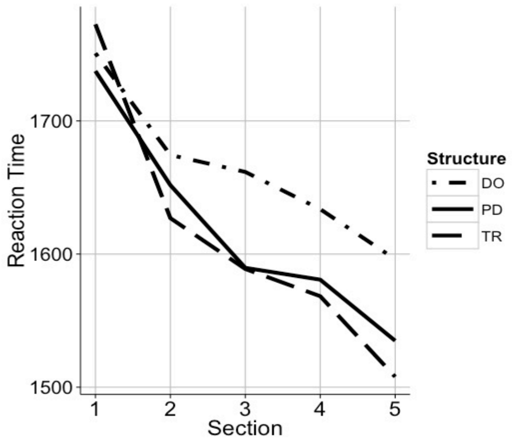
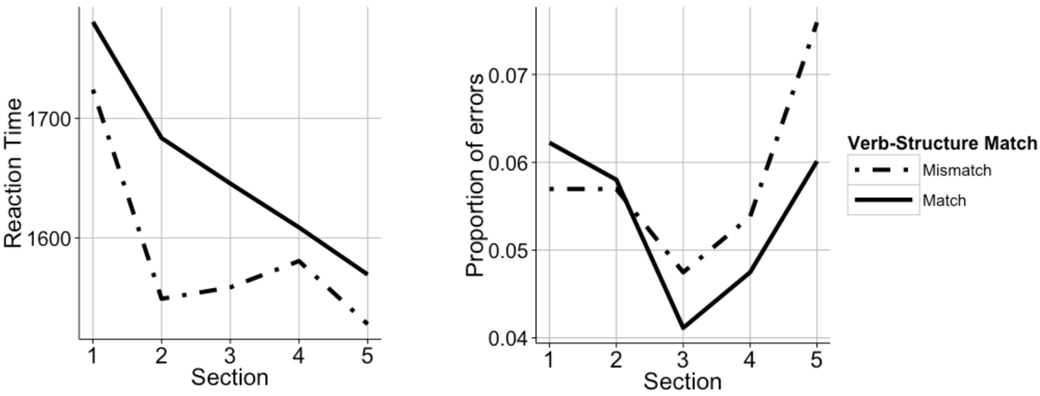


Figure 2.4. Reaction time taken to find INOUN category letters in different structure sequences over the course of the study (section).

2.2.1. Verb bias

The second type of analysis examined the development of verb bias over the study. We examined the reaction time and accuracy on the post-verbal position in PD and DO dative structures. Verb bias analyses included section (centred) crossed with the match between verb's bias and sequence structure, or verb-structure match (match vs. mismatch, effect coded) as predictor variables.

For reaction time, the results showed that people became faster as the experiment progressed, $\beta = -0.03$, $SE = 0.006$, $\chi^2(1) = 25.64$, $p < .001$. There was also a main effect of verb-structure match, where, contrary to our prediction, people were slower to respond when sequence structure matched verb's bias (Fig. 2.5, a), β= 0.04, *SE* = 0.01, $\chi^2(1)$ = 19.4, p < .001. Finally, there was a marginal interaction between section and verb-structure match, showing a trend towards the reaction times dropping faster over section when verb matched sequence structure, β= -0.01, *SE* = 0.007, $\chi^2(1)$ = 3.38, p = .07.



(a) **(b)**

Figure 2.5. Mean reaction times (a) and proportions of errors (b) at different stages in the experiment when verb and structure matched (solid line) or mismatched (dashed line).

Average error rate at the post-verbal position was 5.5% (sd=2.3). Error analysis showed a marginal two-way interaction between section and verb-structure match, meaning that the likelihood of making an error marginally decreased over the course of the experiment when sequences structure matched verb's bias (Figure 2.5, b), β=-0.24, *SE* = 0.11, $\chi^2(1)$ = 3.02, p = .08. No other main effects or interactions were found.

2.2.2. Structural priming

To examine structural priming over development, the accuracy and reaction times for the post-verbal letter in PD and DO target structures was extracted. Mixed models included section (centred) crossed with match between prime and target sequence structure, or prime-target match (match vs. mismatch, effect coded) as predictor variables. The models that converged included random slopes for section crossed with prime-target match. Reaction time analysis revealed that people became faster as the experiment progressed, β= -0.03, *SE* = 0.005, $\chi^2(1) = 36.07$, $p < .001$. There was also a marginal main effect of structural match between prime and target sequences, showing that people responded marginally slower when prime and target structure matched, β= 0.02, *SE* = 0.01, $\chi^2(1) = 3.4$, $p = .07$. No other main effects or interactions were found.

2.3. Experiment 1 Summary

In this study, participants showed improvements in their performance and these improvements were due to both general learning effects and the learning of grammatical regularities. General skill learning can be seen in the improved reaction times over section, while grammar-related knowledge is supported by the INOUN category analysis, which showed that participants required different amounts of time to find the letters of INOUN category depending on which structural context they occurred in. The improvement in speed was accompanied by an increase in error rates, which suggests a speed-accuracy trade-off (Plamondon & Alimi, 1997; Wickelgren, 1977). This trade-off between different measures of sequence knowledge is an important issue that will be addressed in the following 3 studies.

Experiment 1 found that participants were slower to respond when sequence structure matched verb's bias and this is the opposite of what has been found in natural language studies (e.g. Garnsey et al. 1997). However, there was also a marginal interaction between section and verb-structure match in both RT and errors, which indicates a trend towards greater improvements over the course of the experiment when the verb occurred in its preferred structure. This suggests a weak verb bias effect that was growing over the study, and it is possible that this study was too short for participants to fully learn the verb bias in the language.

With respect to structural priming, we found no evidence for such effects in reaction time data. The lack of structural priming effect could be partially attributed to the nature of the task. For example, it has been argued that priming is less robust in comprehension than it is in production (Tooley & Traxler, 2010). Some studies that used reaction time measures found abstract priming using dative structures (e.g., Thothathiri & Snedeker, 2008) while others failed to observe such effects without verb overlap between prime and target sentences (e.g., Arai et al., 2007; Tooley & Bock, 2014). In the current task, participants had to process letter strings one letter at a time, as in comprehension, where sentences reach the listener one word at a time. Thus it is possible that changing the method to make it resemble the process of language production could reveal stronger priming effects.

In general, the weak nature of the verb bias and structural priming effects suggest that participants had trouble learning the structural aspects of the language or did not use that knowledge to facilitate sequence processing. Both verb bias and structural priming depend on the distinction between ANOUN and INOUN category letters, as these categories differentiate the dative DO and PD structures in the post-verbal position. In the present task, the ANOUN and INOUN letters were randomly arranged around the circle, so even if participants could use verb bias or structural knowledge to predict one category, it would be difficult to use that information to increase the speed in selecting the letter. To address this, letters belonging to the same category were clustered together in experiment 2.

3. Experiment 2: Semantic cues and syntactic grouping

To enhance participants' ability to predict and process ANOUN or INOUN letter faster, we replaced the ANOUN and INOUN category letters with images that suggested their animacy. ANOUN letters were replaced with stick-men-like figures, while INOUN letters were replaced with inanimate object-like symbols. Furthermore, letters/symbols belonging to the same category were placed next to each other on the opposite sides of the circle (Figure 4, method). This way people could use their grammatical expectations to anticipate one of another general direction appropriate to PD and DO structures. These changes essentially provided participants with animacy-related categories before they started learning the language. This is appropriate within this task, because our goal is not to examine how syntactic categories are acquired (as in Hunt & Aslin, 2010), but rather to understand how participants learn to use categories within structures. For example, infants seem to distinguish animate and inanimate entities before language production begins (Luo, 2011; Poulin-Dubois, Lepage, & Ferland, 1996; Rakison & Poulin-Dubois, 2001), so it seems likely that by the time that children acquire the dative structures, they have a fairly robust distinction between animates and inanimates. Gropen, Pinker, Hollander, Goldberg, & Wilson (1989) found that in the acquisition of a novel dative verb, children make use of animacy to constrain its use. By providing animate categories through the use of spatial grouping and visual similarity, we are simulating the knowledge the children bring to the learning of the dative alternation.

In terms of the lack of structural priming in the previous study, one possibility was that the effect might have been masked by the comprehension-like nature of the task. To address this, we added a production-like sequence-recall task. On those trials, participants saw a full sequence presented in the centre of a computer screen. Like in comprehension task, participants had to find each letter on the circle. However, when participants moved the mouse cursor from the centre, the sequence disappeared and they had to produce the rest of the sequence from memory. Recall tasks have been used to assess structural knowledge (e.g. Ferreira & Dell, 2000). Using a recall task, Potter and Lombardi (1990; Lombardi & Potter, 1992) presented lure verbs in between exposure to a sentence and its recall. They found that occasionally participants replaced the original verb in the sentence with the lure verb and the structure of the sentence was changed to fit the subcategorisation preferences of the verb. The same method has also been used to test structural priming effects. For example, Potter & Lombardi (1998) found that target sentence recall was affected by the processing of the previous sentence that influenced people to recall the target sentence using the structure of the prime sentence. Thus, a recall task provided a way to look for verb bias and structural priming effects in production using accuracy measures and we also expected to find an effect of verb bias and priming in RTs, as in studies of verb bias (e.g. Jennings et al., 1997) and priming (e.g. Segaert, Menenti, Weber, & Hagoort, 2011; Smith & Wheeldon, 2001; Wheeldon & Smith, 2003).

Finally, since it is possible that the previous experiment was too short for participants to acquire the constraints in our language, we added an additional 24 items to the training set (6 sections), which increased the total number of sequences to 144.

3.1. Method

3.1.1. Participants

An opportunity sample of 38 participants was recruited from the undergraduate student population at the University of Liverpool, who received course credit in exchange for their participation.

3.1.2 Materials

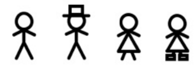
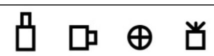
In this version of the circle task, ANOUN category letters were replaced with stickman-like figures, while INOUN category letters were replaced with object-like shapes (Table 2.4). These changes represented animacy cues found in natural language, where stickman-like figures represented animate objects and object-like shapes represented inanimate objects. Category members were also grouped together and presented on the opposite sides of the circle to create visually identifiable groupings to aid category learning. Symbols of the other categories were also grouped together.

The strings of the language were generated in the same way as Experiment 1 (Table 2.4), thereby keeping the verb bias and structural priming manipulation the same. One difference was that IN structure sequences occasionally included an optional prepositional phrase (PREPB INOUN) to counterbalance the more frequent use of the ANOUN category overall. In general, greater exemplar variability is known to facilitate learning of grammatical relationships in artificial grammar learning tasks (Gómez, 2002; Gómez & Maye, 2005).

To give the participants more opportunities to learn the strings, an additional section of 24 items was added increasing the total number of strings to 144.

Table 2.4.

Category types, names, and symbols

Category Type	**Category**	**Symbols**
Animate Noun	ANOUN	
Inanimate Noun	INOUN	
Intransitive Verb	IVERB	W, L
Transitive Verb	TVERB	S, G
Dative verb with PD bias	DVERBP	J, B
Dative verb with DO bias	DVERBD	D, N
Preposition	PREP	C

The sequence-recall task was distributed so that every two target trials in comprehension were followed by two target trials in the sequence-recall task. Two counterbalanced lists were created. In one list, comprehension target trials had DO structure, while sequence-recall target trials had PD structure. All targets had DO-biased verbs and PD-biased verbs 50% of the time. In the second list, all comprehension target trials had PD structure and all sequence-recall trials had DO structure. In both lists, prime sequences were the same and had either DO structure with DO-biased verbs or PD structure with PD-biased verbs.

3.1.3 Procedure

The comprehension task was identical to that in experiment 1, where letter symbols were presented in the centre of the circle one at a time. In the sequence-recall task, the whole string was presented in the centre with the word ‘generate’ to prompt people to remember the string (Fig. 2.6). Once, the mouse moved away from the centre, the string disappeared from the centre and participants were required to complete the sequence from memory. The experiment took approximately 20min to complete.

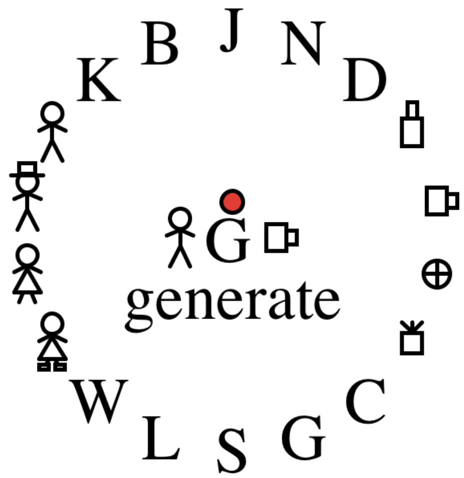


Figure 2.6. Visual display seen by participants in sequence recall task

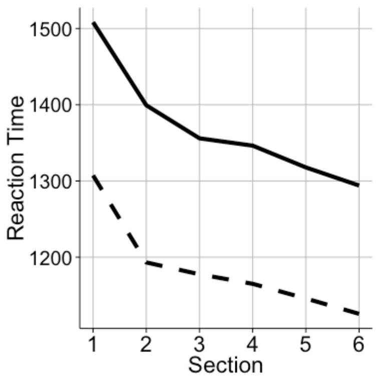
3.1.4 Data Collection and Statistical Analyses

Data were treated and analysed the same way as in Experiment 1. The task produced a total of 22,463 responses, 5.5% of which were incorrect.

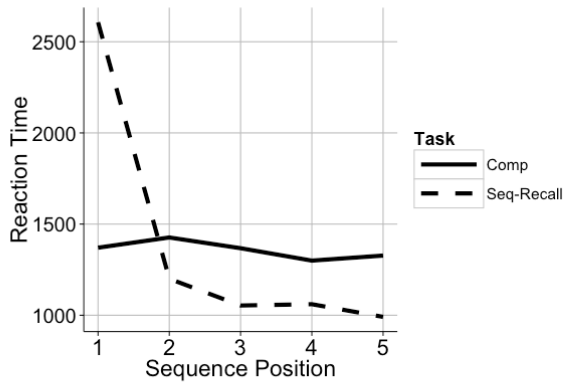
3.2 Results

To examine learning of the grammar, we first tested how participants' reaction times and accuracy changed over the course of the study in different sequence positions in comprehension and sequence-recall. Reaction times were predicted by section (centred) fully crossed with sequence position (centred) and task type (comprehension vs. sequence-recall, effects coded). The maximal model that converged contained random slopes for section, sequence position, task type and two-way interactions between section and task type and between sequence position and task type.

Participants became faster as the experiment progressed, $\beta = -0.03$, $SE = 0.03$, $\chi^2(1) = 7.52$, $p = .006$. They were faster in later positions of the sequences, showing that greater amount of structural context helped them find letters more effectively, $\beta = -0.07$, $SE = 0.005$, $\chi^2(1) = 8.16$, $p = .004$. They were faster overall in the sequence-recall task $\beta = -0.15$, $SE = 0.01$, $\chi^2(1) = 47.54$, $p < .001$. This is because in the sequence-recall participants did not have to wait to find out which letters constituted the sequence. On average, increase in speed in later positions was greater in the sequence-recall task, $\beta = -0.11$, $SE = 0.01$, $\chi^2(1) = 49.36$, $p < .001$. Also, the two-way interaction between section and sequence position showed that the overall increase in speed in later sequence positions became more prominent as the experiment progressed, $\beta = -0.01$, $SE = 0.01$, $\chi^2(1) = 19.41$, $p < .001$. This increase was also greater in the sequence-recall task, $\beta = -0.02$, $SE = 0.003$, $\chi^2(1) = 33.75$, $p < .001$.



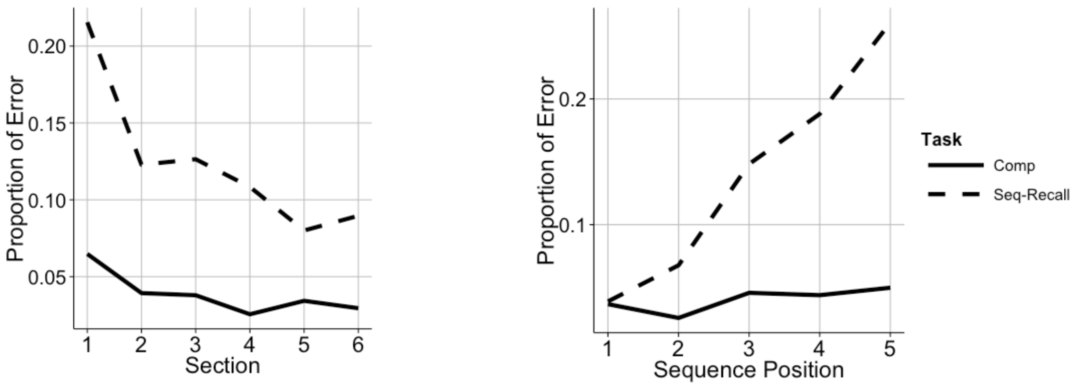
(a)



(b)

Figure 2.7. Average reaction time in different sections (a) and at different sequence positions (b) in comprehension (solid line) and sequence-recall task (dashed line).

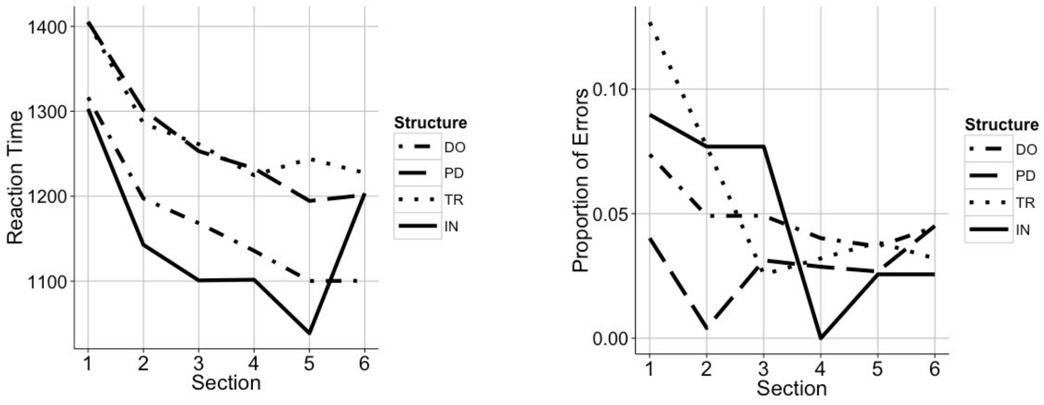
Accuracy analysis showed that participants began making fewer mistakes as the experiment progressed, β= -0.25, *SE* = 0.03, $\chi^2(1) = 6.46$, $p = .01$. Later sequence positions were associated with more mistakes than earlier sequence positions, β= 0.34, *SE* = 0.03, $\chi^2(1) = 13.63$, $p < .001$. On average, participants also made more errors in sequence-recall task than in comprehension, β= 1.1, *SE* = 0.12, $\chi^2(1) = 74.38$, $p < .001$. This reflected greater task difficulty in the sequence-recall task due to the memory component. Finally, a two-way interaction showed that participants made more errors in later sequence positions in the sequence-recall task than in comprehension, β= 0.45, *SE* = 0.03, $\chi^2(1) = 80.16$, $p < .001$.



(a) **(b)**

Figure 2.8. Average proportion of errors in different sections (a) and at different sequence positions (b) in comprehension (solid line) and sequence-recall task (dashed line).

Reaction time and accuracy analyses were then performed on INOUN category letters in different structures to ensure that the participants were learning grammatical knowledge. This analysis was performed on comprehension trials only because IN and TR structure sequences never occurred in the sequence-recall task. Reaction times were predicted with section crossed with structure (as factor). The model included a random slope for structure. The results showed that on average, participants became faster at finding INOUN category letters as the experiment progressed (Fig. 2.9, a), $\chi^2(1) = 239$, $p < .001$. There was also an effect of structure, as the speed with which participants were able to find INOUN category letters differed between structures, $\chi^2(1) = 61.94$ (3), $p < .001$. Finally, there was an interaction between section and structure, $\chi^2(1) = 9.03$ (3), $p = .03$. This shows that with the progression of the experiment, people became increasingly sensitive to structural contexts in which INOUN category occurred.



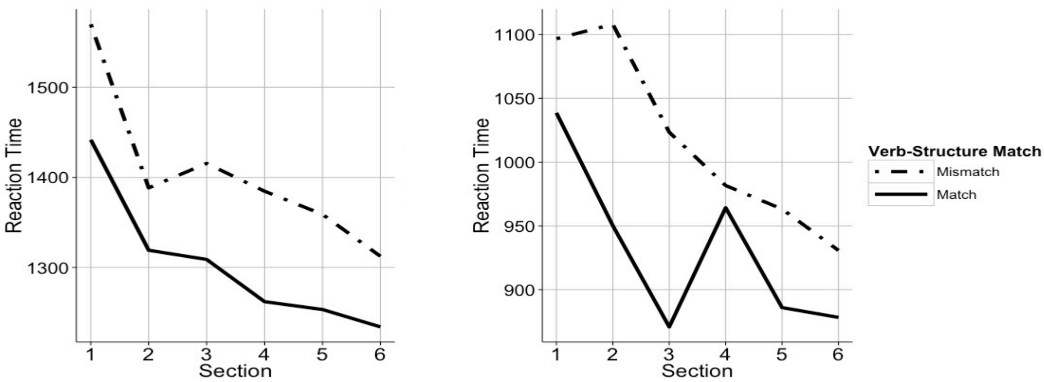
(a) **(b)**

Figure 2.9. Reaction times (a) and error rates (b) when selecting INOUN category letters in different structure sequences over the course of the study.

For the accuracy analysis, a mixed model with the same predictor variables and a random intercept showed that the average error rate dropped as the experiment progressed (2.9, b), $\chi^2(1)$ = 11.25, p < .001. There was a main effect of structure, showing that participants' error ratings were different for different rules, $\chi^2(1)$ = 11.54 (3), p < .001. Finally, a two way interaction between section and rules showed that over the course of the study, accuracy improved at different rates for different rules, $\chi^2(1)$ = 14.2, p < .003. In combination with the reaction time analysis, these findings show that participants were not just becoming faster overall, but were learning position and grammatical expectations that changed the speed and accuracy of their performance.

3.2.1 Verb Bias and Reaction Times

To test the effect of verb bias and processing speed, reaction times were analysed separately for comprehension and sequence-recall tasks. In comprehension, a mixed effects model with section (centred) crossed with verb-structure match and random slopes for the same variables revealed a main effect of section, showing that participants' reaction times decreased with the progress of the experiment (Fig 2.10, a), β= -0.03, *SE* = 0.003, $\chi^2(1)$=28.67, p <0.001. The results also showed a main effect of verb-structure match, meaning that people were faster when sequence structure matched verb's bias, β= -0.08, *SE* = 0.019, $\chi^2(1)$=28.53, p <0.001.



(a) **(b)**

Figure 2.10. Reaction time in comprehension (a) and sequence-recall (b) tasks when verbs occurred in their preferred (solid line) or non-preferred (dashed line) structure

The same analysis was performed on sequence-recall trials. Consistent with comprehension trials, the results showed that participants became increasingly faster as the experiment progressed (Fig. 2.10, b), β= -0.03, *SE* = 0.007, χ^2= 14.3, p < .001. Participants were also faster when verb matched its structure, showing verb bias effects, β= -0.1, *SE* = 0.05, χ^2 = 13.9, p < .001.

3.2.2 Verb bias and accuracy

Average error rate at the post verbal position was 2.4% (sd=1.5) in the comprehension task and 6.2% (sd=2.4) in the sequence-recall task. A mixed effect with section crossed with verb-structure match and random slopes for the same variables showed a marginal main effect of verb-structure match, suggesting that people made marginally more errors when the verb matched its structure, β= 0.95, *SE* = 0.36, $\chi^2(1) = 3.38$, $p = .07$. No other main effects or interactions were observed.

The same analysis was then performed on the sequence-recall trials. The results revealed a significant main effect of section, showing that error rates went down as the experiment progressed, β= -0.33, *SE* = 0.08, $\chi^2(1) = 13.61$, $p < .001$. A main effect of verb-structure match was also observed, showing that people were less accurate when verb bias matched its structure (Fig. 2.11), β= 0.59, *SE* = 0.22, $\chi^2(1) = 5.53$, $p = .02$. No other main effects or interactions were found.

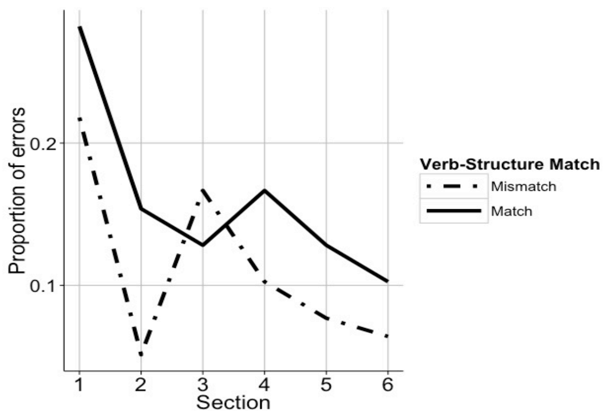


Figure 2.11. Proportion of error in sequence recall task over section by verb-structure match

3.2.3 Structural priming and reaction times

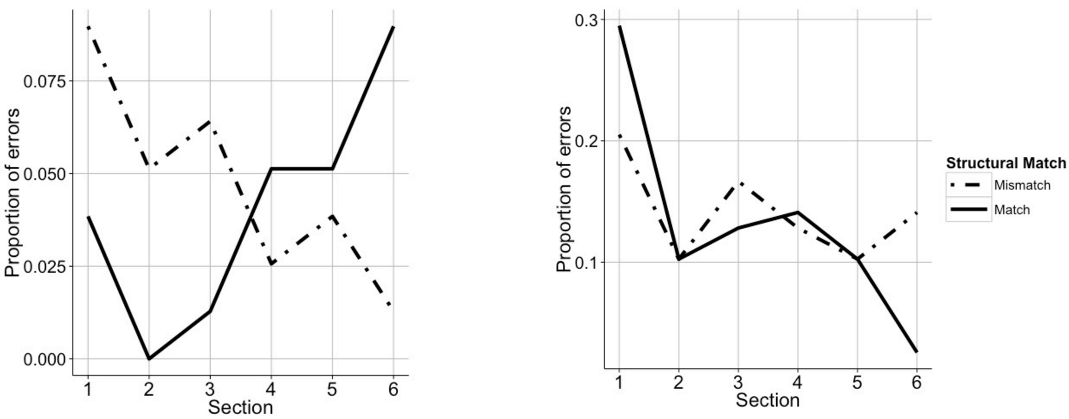
The influence of structural priming on processing speed in comprehension was tested by predicting reaction times with section (centred) crossed with prime-target match. The model that converged included random slopes for section and structural match. A significant main effect of section showed that people became faster as the experiment progressed, β= -0.03, *SE* = 0.004, $\chi^2(1) = 43.37$, $p < .001$. No other effects were observed.

The same analysis was then performed on the sequence-recall task. Like in comprehension, people became faster as the experiment progressed, β= -0.03, *SE* = 0.007, $\chi^2(1) = 17.06$, $p < .001$. No other effects were found.

3.2.4 Structural priming and accuracy in Sequence-Recall task

Mean error rate at the post-verbal position in the sequence-recall task was 13.7% (sd=3.43). To test structural priming effects, error rates were predicted with section crossed with structural match. Mixed effects structure included random slopes for the same variables. Participants' errors dropped as the experiment progressed, showing practice effects, β= -0.3, *SE* = 0.23, $\chi^2(1) = 14.07$, $p < .001$. The results also showed that as the experiment progressed, people began producing marginally fewer mistakes when target structure matched prime structure, suggesting growing priming effects, β= -0.3, *SE* = 0.15, $\chi^2(1) = 2.89$, $p = .09$.

To explore how the growing priming effect in the sequence-recall task compared to the comprehension task, we repeated the same analysis on combined accuracy data. The model included task type as an additional predictor variable that was fully crossed with section and prime-target match. The random effects structure included random slopes for section, prime-target match and task type. A general learning effect was indicated by a main effect of section, showing that error rates decreased as the experiment progressed, β= -0.17, *SE* = 0.07, $\chi^2(1) = 8.1$, $p = .004$. There was also a main effect of task type, showing that participants produced more errors in sequence recall task, relative to comprehension, β= 1.3, *SE* = 0.22, $\chi^2(1) = 30$, $p < .001$. These differences reflected task demands. There was a marginal two-way interaction between section and task type, where people improved marginally more in the sequence recall task relative to comprehension, β= -0.25, *SE* = 0.12, $\chi^2(1) = 3.6$, $p = .06$. Finally, we found a significant three-way interaction between item, structural match and task type, β= -0.97, *SE* = 0.24, $\chi^2(1) = 15.05$, $p < .001$. These results showed that people improved significantly more over the course of the study when prime and target structures matched in the sequence-recall task relative to comprehension, showing a growing priming effect in production (Fig. 2.12). No other main effects or interactions were found.



(a) **(b)**

Figure 2.12. Proportion of errors when finding the post verbal letter in sequences that were primed with the same (solid line) or different (dashed line) structure sequence in comprehension (a) and sequence-recall (b) task

3.3. Experiment 2 Summary

Like in experiment 1, we found evidence that participants were not just learning task-specific abilities (like moving the mouse to particular letters), but that they were encoding the regularities in the grammar. In contrast to the experiment 1, we found verb bias effects in reaction times both in comprehension and recall tasks.

This is consistent with natural language studies that found faster processing times in comprehension (Garnsey et al., 1997) and production (Jennings et al., 1997) when sequence structure was consistent with verbs' biases. The fact that verb bias effects were found in the present study and not the previous study suggests that the grouping of the symbols and animacy cues enhanced the ability to exhibit verb bias knowledge.

It is worth noting the different direction of verb-structure match effect in accuracy measures. This effect was marginal in comprehension but significant in the sequence-recall task. While these effects seem contradictory, the opposite nature of these results suggest that the two may be inextricably linked and reflect the same underlying processes. Rapid human behaviours are characterized by speed-accuracy trade-offs in both non-linguistic (Plamondon & Alimi, 1997; Wickelgren, 1977) and linguistic tasks (MacKay, 1982). That is, faster performance is associated with greater likelihood of making an error and thus these two measures should trade off against each other. To test if this could explain the direction of the accuracy and reaction time results, a separate analysis was performed where errors were submitted to a mixed-effect logistic regression with reaction times as a predictor variable. The results revealed that reaction time was a significant predictor of error rates, showing that slower reaction time was associated with lower error rate (β= -0.001, SE = 0.0005, $x^2 = 10.32$, $p = .001$). It may be the case that task demands can cause variation in the speed/accuracy trade-off and this could vary across and within participants.

In terms of structural priming, participants became increasingly more accurate at finding the post verbal letter in the sequence-recall task relative to the comprehension task when target sequences were primed with the same sequence structure. This is the first evidence that structural priming-like effects can occur in a non-linguistic sequence learning task. As shown by the grammar learning analysis, the sequence-recall task was associated with a larger overall learning effect, where later sequence positions were associated with greater increase in speed. The greater sensitivity of the sequence-recall task could explain why only this task showed structural priming effects.

In sum, the first two studies provide evidence that both verb bias and structural priming effects occur in the Circle task. This was achieved with the help of grouping the same category letters next to each other on the circle and incorporating visual shape cues that facilitated learning the distinctions between INOUN and ANOUN category letters. However, since the structural priming effect was a relatively small, it is important to replicate this effect. Furthermore, the use of visual shapes to encode animacy may have rendered the processing of the ANOUN and INOUN symbols less automatic than the processing of the other symbols, and it would be prudent to know whether the same results are possible using noun letters that are similar to the letters used for verbs and prepositions.

4. Experiment 3: Highlighted letters

According to Abrahamse, Jiménez, Verwey, & Clegg (2010) there are many features that people may be learning in serial response time tasks. These can include stimulus-dependent features like learning associations between stimulus locations, associations between shapes (e.g. letters) or between shapes and locations, or response-dependent features where people learn the directions and patterns of movement or associations between stimulus and response features. In eye-tracking studies, participants show anticipatory eye movements by looking at the objects that are most compatible with previous linguistic input (Altmann & Kamide, 1999). Since the Circle task requires participants to focus on the central location of the circle where the stimulus letters were presented, they had less freedom to move their eyes to anticipate upcoming symbols in the language. To reduce this constraint, we presented letter sequences by highlighting the letters one at a time on the circle of letters by changing their color to green temporarily. As in the previous two studies the task was to move the mouse to the target letter as fast and accurately as possible. In this version of the task, if participants could use their grammatical knowledge to look at the right part of the circle, they would be faster to respond to letters that appeared on that part of the circle and this might make the task more sensitive to grammatical knowledge.

Furthermore, the sequence recall task was changed to a sentence production task. Experiment 1 did not find an effect of structural priming and experiment 2 found only a weak effect in a sequence-recall task. This could be due to the fact that the tasks differed from human priming studies in several important ways. The majority of structural priming studies in natural language use sentence production methods that measure structural choice tendencies. For example, Pickering & Branigan (1998) presented participants with a sentence fragment that contained a noun and a verb (e.g. The patient showed…) and asked them to complete the sentence by creating their own ending. They found that participants showed robust priming in the sentences that they produced. Sentence completion tasks are also used to measure verb bias (Garnsey et al., 1997; Trueswell et al., 1993). To make our task closer to these tasks, we replaced our sequence-recall task with a sequence-completion task, where people were presented with the first two letters of the string (equivalent to noun and verb) in the centre without highlighting those letters on the circle. Those letters disappeared after a second and people had to find them on the circle. After this, two possible letters were highlighted. One letter belonged to the ANOUN category, while the other one belonged to the INOUN category. People had to select whichever letter they preferred on that occasion and their choice determined the structure of the sequence. This allowed us to measure the proportion of selected structures based on verbs' biases and the structure that the sequence was primed with.

The hypothesis for the comprehension task remained the same as in the first two experiments. It was predicted that participants would be faster and more accurate at finding the first post verbal letter when sequence structures matched verb's bias. For the sequence-completion task, it was hypothesized that people would produce more structures that are consistent with verb's bias and that they would also be faster and more accurate at doing so. In terms of structural priming, it was hypothesized that people would be faster at finding the post verbal letter of the target sequence that was primed with the same structure sequence. It was also expected that participants would produce more structures in a sequence-completion task that were consistent with that of a prime sequence, and that they would be faster and more accurate when their chosen structure matched the structure of the prime.

4.1. Method

4.1.1. Participants

70 participants were recruited from the undergraduate student population at the University of Liverpool and received course credit in exchange for their time.

4.1.2. Materials

The strings of the language were generated the same way as in the previous experiment, by randomly selecting letters appropriate for the category (Table 2.5).

Verb bias was created by presenting DVERBP and DVERBD category letters 66% of the time in PD and DO structures respectively. The proportion was slightly smaller as in than in the previous two experiments (75%) because there was no control over what structure participants would produce in the sequence-completion task, and thus only comprehension trials could be used to manipulate the frequencies with which verbs occurred in different structures. Context for structural priming was the same way as in the previous two experiments by presenting PD and DO structures in pairs in all combinations (PD-PD, DO-DO, PD-DO, DO-PD), where one sequence acted as a prime and the other one was a target. Each pair was separated by a filler sequence that had either transitive or intransitive structure. Like in the previous experiment, each pair occurred twice within each section. Half of the target sequences were used to test structural priming in comprehension, while the other half was used for the sequence-completion task. Six such sections containing 8 prime-target pairs were created, which produced a training set of 144 items. Two counterbalanced lists were created in the same way as in Experiment 2.

Table 2.5.

Category types, names, and symbols

Category Type	**Category**	**Symbols**
Animate Noun	ANOUN	X, M, Y, H
Inanimate Noun	INOUN	F, Z, Q, P
Intransitive Verb	IVERB	L, S
Transitive Verb	TVERB	G, C
Dative verb with PD bias	DVERBP	B, K

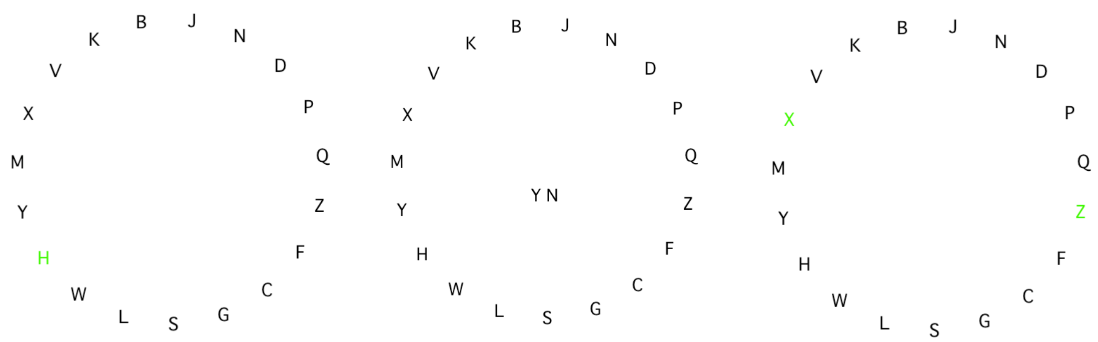
Dative verb with DO bias	DVERBD	N, J
Preposition 1	PREP	D
Preposition 2	PREP2	W
Adverb (only occurred in sentence-completion task when participants selected DO structure)	ADV	V

4.1.3. Procedure

The experiment was run using the same computer systems and software as in experiment 2. The visual display consisted of letter symbols forming a circle (Fig. 2.13, a). Letters belonging to the same category were presented on the circle adjacent to each other. ANOUN and INOUN categories were placed on the opposite sides of the circle.

In the comprehension task, strings were presented by highlighting the appropriate letters on the circle at a time by changing the colour to green. The task was to move the mouse cursor on top of the highlighted letter as fast and as accurately as possible. Each response reset the mouse back to the centre and triggered the appearance of the next letter. A short delay of 200ms was added before the next letter was highlighted to allow participants some time to anticipate the direction of the next possible location. Different letter strings were separated by a blank screen where participants could press the spacebar to continue the experiment.

In the sequence-completion task, the first two letters of the sequence were presented in the centre of the circle (Fig. 2.13, b) with no letters highlighted on the circle. The letters disappeared after 2 seconds, after which participants had to find those letters on the circle. Then two letters were highlighted (Fig. 2.13, c), one of which belonged to the ANOUN, while the other belonged to the INOUN category. People had to choose only one of the highlighted letters. If participants selected the ANOUN category letter, the sequence continued as a DO structure sequence, as in comprehension task. If the choice was the INOUN category letter, then the sentence continued as a PD structure sequence. To prevent participants from learning that their choice corresponding to the DO structure meant a shorter sequence, one additional letter (V) was added to the circle that completed the DO sequence. This letter played a role of an adverb (e.g. boys give girls books slowly) and occurred only in the sentence completion task when participants selected DO continuation. This was done to ensure that people's choices were not influenced by the differences in the length of the two structures.



(a) **(b)** **(c)**

Figure 2.13. The visual display seen by participants in comprehension (a) and sequence completion task where people saw the first two letters (b) and then had to choose from the two highlighted letters (c) to continue the sequence

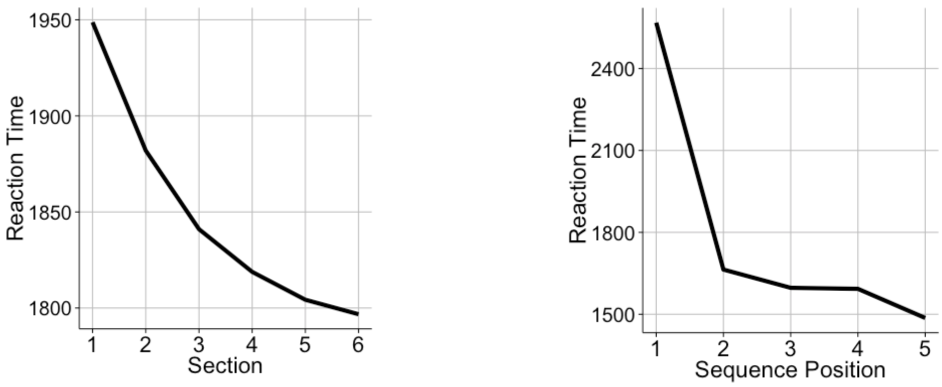
Like in the previous experiment, up to 6 participants were tested on separate computers in a quiet room. Participants read instructions on a computer screen that informed them that they would have to process letter sequences of various lengths by finding their location on the circle of letters using a mouse cursor when the letters were highlighted. The goal was to complete the task as fast and as accurately as possible. Participants were not told that the letter sequences followed certain rules. To explain their task on sequence-completion trials, the participants were told that once they had to make a choice between two highlighted letters, they could choose whichever letter they preferred on that occasion. The whole experiment took approximately 20 minutes to complete.

4.2. Results

Statistical analyses followed the same assumptions as in experiments 1 and 2. Comprehension trials produced 32781 responses with error rate of 5.3%. Sequence-completion trials produced 7926 responses with error rate of 3.9%. In the sequence-completion task, proportions of produced PD and DO structures were calculated.

The first analysis tested whether participants were learning grammatical constraints. Only comprehension trials were analysed because sequence completion involved different tasks at different word positions. Reaction times and error rates were predicted by section crossed with sequence position in separate analyses. Maximal models included random slopes for section crossed with sequence position.

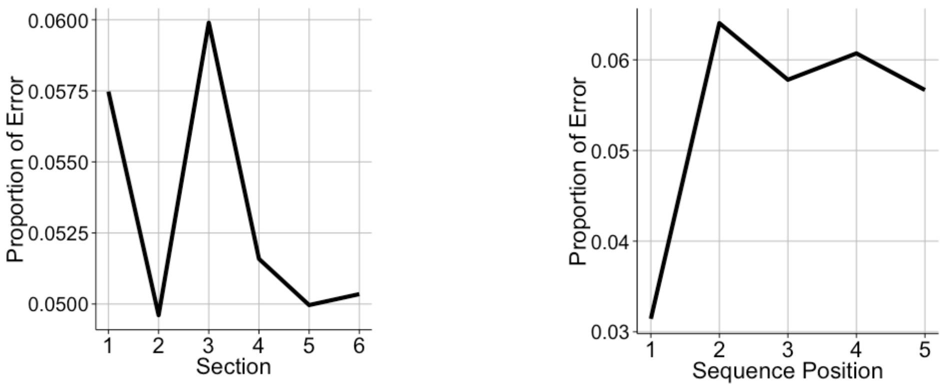
Despite the trend seen in Figure 14 (a), the reaction time analysis showed no significant increase in speed over the course of the study. However, it is seen in Figure 14 (b) that the changes were evident in later sequence positions where participants' speed increased, as compared to earlier sequence positions, $\beta = -0.13$, $SE = 0.002$, $\chi^2(1) = 265.41$, $p < .001$. However, the increase in speed became smaller as the study progressed, which shows that participants were learning position-specific processing biases, $\beta = 0.003$, $SE = 0.001$, $\chi^2(1) = 30.73$, $p < .001$.



(a) **(b)**

Figure 2.14. Average reaction time in different sections (a) and at different sequence positions (b) in comprehension

Accuracy analysis showed that participants' accuracy improved as the experiment progressed,β= -0.07, *SE* = 0.03, $\chi^2(1)$= 9.5, *p* = .002. Participants showed position-specific biases, where accuracy was reduced later in sequences, β= 0.16, *SE* = 0.03, $\chi^2(1)$= 27.1, *p* < .001.

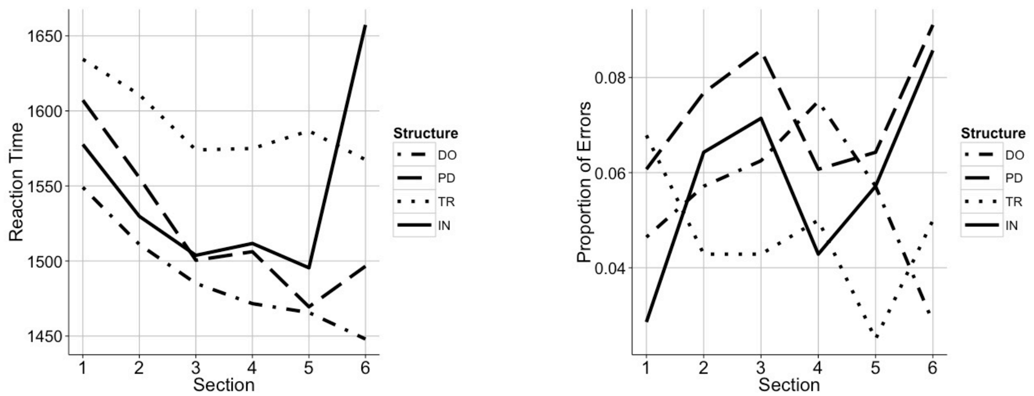


(a) **(b)**

Figure 2.15. Average proportion of errors in different sections (a) and at different sequence positions (b) in comprehension.

Performance was then tested on INOUN category letters in different structures over the course of the study. Reaction times and error rates were predicted by section crossed with structure (as factor) in separate analyses. The maximal models that converged included random slopes for section and structure in reaction time analysis and only the intercept in the accuracy data analysis.

Reaction times dropped overall for INOUN category letters over the course of the study, $\chi^2(1)$= 16.9 (3), $p < .001$. There was a main effect of structure, showing that participants responded to INOUN category letters in different structures at different speeds, $\chi^2(1)$= 70.32 (3), $p < .001$. Section also interacted with structure type, showing that reaction times dropped for different structures at different rates, $\chi^2(1)$= 34.32 (3), $p < .001$.



(a) **(b)**

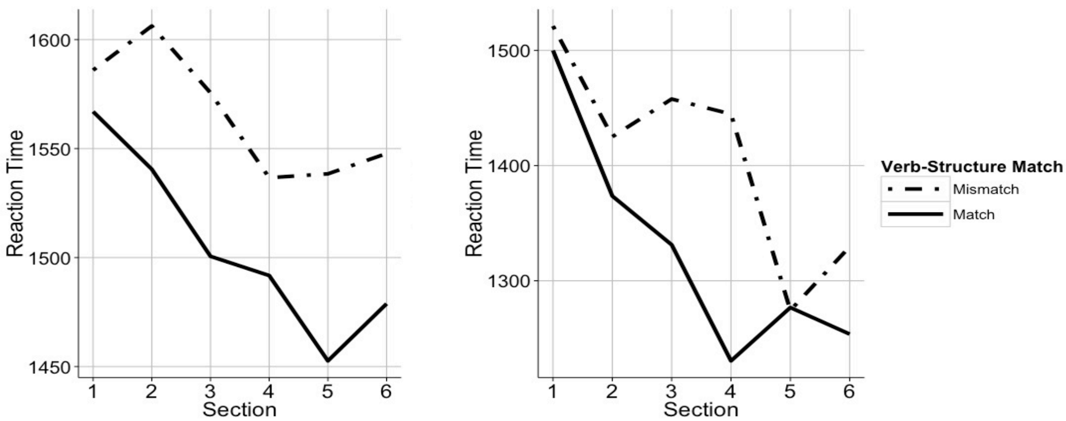
Figure 2.16. Reaction times (a) and error rates (b) when selecting INOUN categoryletters in different structure sequences over the course of the study.

An error analysis with the same predictor variables and random structure showed that participants produced different proportions of errors for different structures, $\chi^2(1)$= 16.16 (3), p = .001). No other effects or interactions were found.

Thus, as in the previous studies, participants are learning grammar-specific knowledge and not just showing general practical effects.

4.2.1 Verb bias and reaction time

To test the effects of verb bias on comprehension speed, reaction times at the post verbal letter position were predicted by section (centred) crossed with verb-structure match. The maximal model that converged contained random slopes for the same two variables. The results showed that people became faster as the experiment progressed, showing a general learning effect, β = -0.01, *SE* = 0.002, $\chi^2(1)$ = 33.73, $p < .001$. There was also a main effect of verb-structure match showing that people were faster when sequence structure matched verb's bias (Fig. 2.17, a), β = -0.03, *SE* = 0.01, $\chi^2(1)$ = 26.28, $p < .001$.



(a) **(b)**

Figure 2.17. Mean reaction time at different sections in comprehension (a) and sequence completion (b) tasks when verb occurred in it's matching structure or when participants chose the structure consistent with the verb's bias in the sequence-completion task respectively.

The same analysis of the sequence-completion trials revealed that participants began making their choices faster as the experiment progressed, indicative of practice effects, β= -0.03, *SE* = 0.005, $\chi^2(1) = 35.78$, $p < .001$. Like in comprehension, they were also faster when making a choice that was consistent with verb's bias (Fig. 2.17, b), β= -0.04, *SE* = 0.013, $\chi^2(1) = 8.35$, $p = .004$. No other effects or interactions were found.

4.2.2. Verb bias and accuracy

Accuracy at the post verbal position was 6.2% (SD = 2.4) in comprehension and 5.8% (sd=2.3) in the sequence-completion task. Only random intercepts were included in the random structure of the model. The comprehension results showed a two way interaction between section and verb-structure match, indicating that over the course of the study participants showed greater improvement in accuracy when sequence structure matched verb's structural preference (Fig. 2.18), β= -0.21, *SE* = 0.1, $\chi^2(1) = 4.37$, $p = .04$. No other effects were observed.

Analysis of error rates in the sequence-completion task revealed no significant effects or interactions.

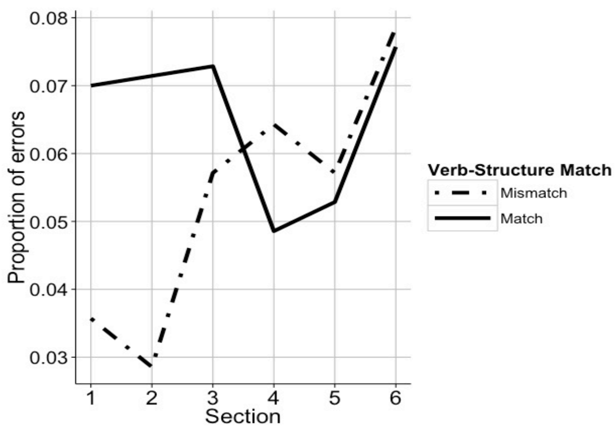


Figure 2.18. Mean error rates in different sections in comprehension task when verb occurred in its preferred (solid line) or non-preferred structure (dashed line)

4.2.3. Verb bias and structural choice

The effect of verb bias on structural choice was analysed by predicting the proportion of produced PD structures with section crossed with verb type (DVERBD or DVERBP, effect coded). The model that converged contained a random slope for section. The results revealed a significant effect of verb type, showing that participants produced more PD structure sequences when the choice was made following PD structure biased verbs (Fig. 2.19), $\beta = 0.29$, $SE = 0.11$, $\chi^2(1) = 6.68$, $p = .01$.

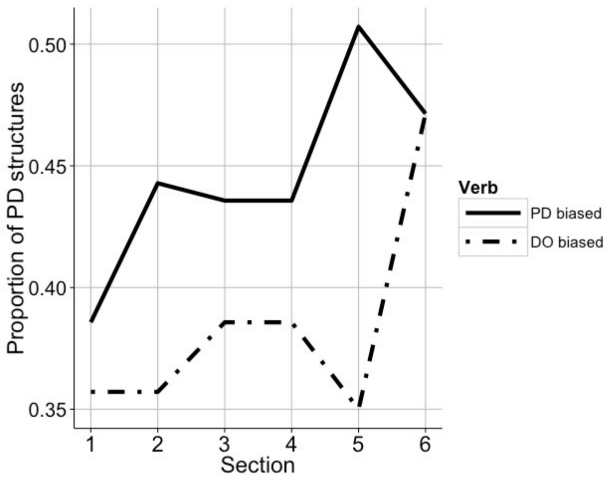


Figure 2.19. Proportion of PD structure choices given PD-biased verb (solid line) and DO-biased verb (dashed line)

4.2.4 Structural priming and reaction times

The influence of structural priming on processing speed in comprehension was tested in a model predicting reaction times with section crossed with prime-target match. The model included random slopes for section crossed with structural match. The results revealed a significant main effect of section, showing that people became faster as the experiment progressed, which indicates a general practice effect, β= -0.01, *SE* = 0.002, $\chi^2(1) = 18.3$, $p < .001$. No other main effects or interactions were found.

The same analysis was then performed on the sequence-completion trials, which also showed that participants became faster at selecting the post-verbal letter, indicating a general practice effect, β= -0.03, *SE* = 0.005, $\chi^2(1) = 37.87$, $p < .001$. No other main effects or interactions were found.

4.2.5 Structural priming and accuracy

To explore the effect of structural priming on the accuracy in the sequence-completion task, sequence-completion errors were predicted by section crossed with prime-target match. The maximal model that converged contained random slopes for section. The results revealed a significant effect of structural match, where participants made fewer errors when they chose the structure that matched the prime sequence structure (Fig. 2.20), β= -1.16, *SE* = 0.26, $\chi^2(1) = 24.42$, $p < .001$. This provides evidence that structural priming occurred in this task in terms of structural choice.

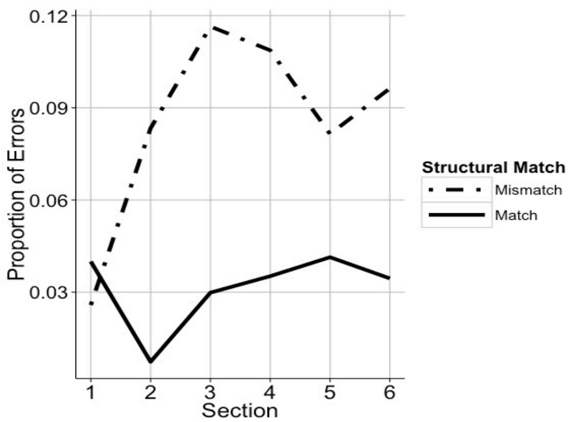


Figure 2.20. Mean proportion of error in different sections in the sequence-completion task when participants chose the structure that was consistent (solid line) or inconsistent (dashed line) with prime sequence structure.

4.2.6. Structural priming and structural choice

To test the influence of the prime structure on participants' choice of the target structure in sequence completion task, the proportions of produced PD structures were predicted by section crossed with prime sequence structure (PD or DO, effects coded). The model included random slopes for section crossed with prime structure.

The results showed a marginal effect of section, which indicates a trend towards people producing more PD structures as the experiment progressed (β= 0.07, *SE* = 0.04, $\chi^2(1)$= 3.1, *p* = .08). Importantly, there was a significant two-way interaction between section and prime sequence structure, showing that with the progress of the experiment people began producing more PD structure sequences following PD primes showing a growing priming effect (β= 0.12, *SE* = 0.06, $\chi^2(1)$ = 3.75, *p* = .05).

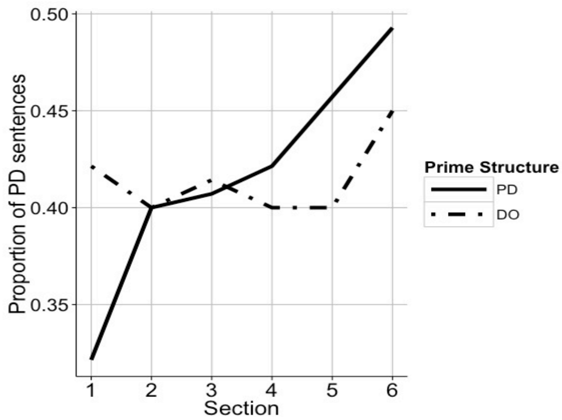


Figure 2.21. Proportion of sequence completed using PD structure after PD (solid line) or DO (dashed line) structure prime sequences over section.

4.3. Experiment 3 Summary

As in the previous studies, participants exhibited position and grammatical biases, showing that they acquired the grammatical aspects of the languages that they were exposed to. The study found verb bias effects in reaction times in the comprehension task. Despite the changes in the way the letter strings were presented, learning verb bias constraints remained robust across the tasks. Like in the sequence-recall task in Experiment 2, verb bias effects were found in reaction times in the production-like sequence-completion task. Participants were faster at making their choice if they chose the post verbal letter that belonged to the preferred category of the verb considering its structural bias. It is important to note that the sequence-completion task always involved free choice of the structure, so these effects show a transfer of the knowledge acquired during the comprehension-like task. This is further confirmed by the structural choice measure, where participants produced a greater proportion of PD structures given PD-biased verbs. This is consistent with natural language studies, where verb bias was found to increase sentence processing speed (Jennings et al., 1997) and structural choices (Garnsey et al., 1997; Trueswell et al., 1993).

As in the previous studies, no structural priming effects were found in the comprehension task. Like in experiment 2, the present experiment showed structural priming in sequence-completion accuracy. People made fewer mistakes when selecting the letter that led to the structure that was consistent with that of the prime sequence, which was evident early in the experiment. In the previous experiment, where only the sequence-recall task was analysed separately from comprehension, the growing structural priming effects were marginal. The method of highlighting the letters instead of presenting them in the centre enhanced this effect. This is likely because people could use the acquired knowledge to anticipate the spatial location of the letter and find the predicted letter easier in comparison to concentrating on the centre of the circle where the stimulus letter appeared, before looking for the spatial location of that letter. Thus it seems that this task was a more sensitive method to study structural priming.

Importantly, the present experiment also found a priming effect in structural choice measure, where participants produced an increasingly greater proportion of PD structures following PD structure primes as the experiment progressed. This means that not only does the processing of the prime affect people's expectations in the comprehension task, but it also translated into their own choices in the sequence-completion task. This mimics natural language studies, where people were shown to favour prime sequence structure when asked to complete fragments of sequences (e.g. Pickering & Branigan, 1998).

Since language-like effects occur in these non-linguistic tasks, an interesting question is the extent to which the learning mechanism is similar between these tasks. Natural language studies have found that the strength of structural priming is affected by the verb's bias in the prime sentence (Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2007). Priming is stronger when the structure of the prime sentence is inconsistent with the verb's bias. That is, if a DO structure prime contains a PD-biased verb, priming is likely to be stronger making people even more likely to produce a DO structure sentence than if the DO sentence contained DO-biased verb. This verb surprisal effect is an important prediction of error-based learning accounts of structural priming (Chang, et al. 2006), because mismatching verb-structure pairings should create more error.

The present study was not ideal for examining whether structural priming in this task is error-driven, because all prime sequences had verbs with matching structural preferences. In addition, one potential problem in the current task is that participants' structural choice in the sequence-completion task is affected by the structural preference of the verb used in the target sequence. Structural priming can compete with verb bias effects when participants make their choice, which adds variability to their responses. Considering that participants are tested in the sequence-completion task only 4 times in each section, that variability may make it difficult to capture subtle effects of the prime verb-structure match on priming. Thus the verbs in the target sentences were replaced with novel verb letters that had no bias towards either structure. The following study then manipulated the match between structure and verb's bias in the prime sequences looking at the proportions of the produced PD structure sequences in a sequence-completion task. Half of the prime sequences contained the verbs that had a bias toward that structure, while the other half of the sequences had mismatching verbs. Based on the natural language studies, it was predicted that people would produce more structures consistent with the prime if the verbs of the prime sequences had a bias towards an alternative structure.

Finally, the previous studies showed variability across different counterbalancing lists due to the different order of items in each list. This variability is difficult to capture with random slopes/intercepts, because each individual learns at different non-linear rates and it is hard to separate out individual versus counterbalancing list variation in our models. Therefore, the final study used only one counterbalancing list to reduce list variability to increase the chance of finding structural priming effects in comprehension.

5. Experiment 4: Prediction-based learning

5.1. Method

5.1.1. Participants

36 participants were recruited from the University of Liverpool student population, who were awarded course credits for taking part.

5.1.2. Materials

The grammar was generated the same way as in Experiment 3. To create the verb bias, 66% of DVERBD verbs occurred in DO and vice versa for DVERBP and

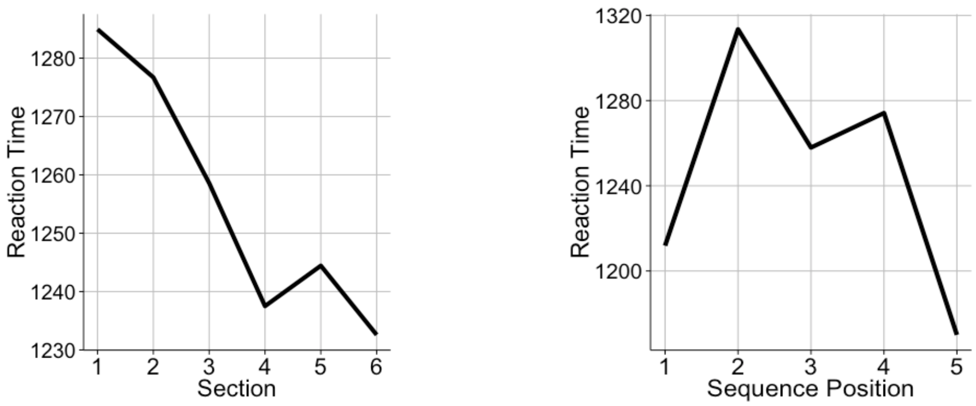
PD structures. Since the focus of the study was on the effect of verb-structure mismatch in the prime sequence on the structural choice of the target sequence, an additional novel dative verb (letter T) was introduced in the sequence-completion task that had no bias toward PD or DO structure. This was done to ensure that participant's response was not influenced by the structural preference of the verb and to make the measure more sensitive to error-based changes related to verb structure pairing in the prime sequences. Since all sequence completion trials had the same verb, structural choices would not be strongly biased by the target verb. Unlike the previous experiment, sequence completion trials were the last 4 target sequences in each section. This ensured that participants had more exposure to the language before sequence-completion trials. To manipulate verb-structure pairing of the prime sequences, sequence completions trials in each section were preceded by every verb-structure (PD-DVERBD, PD-DVERBP, DO-DVERBD, DO-DVERBP) combination once.

The procedure and data treatment and analyses assumptions were the same as in Experiment 3.

5.2. Results

A total of 20580 responses were collected, with an error rate of 5%. As in the previous experiment, grammar learning was analysed from reaction times and error rate in different sections and different sequence positions in comprehension. Both dependent variables were predicted by section (centred) crossed with sequence position in mixed effects linear regression and mixed effects logistic regression respectively. Both models contained random slopes for the same two variables.

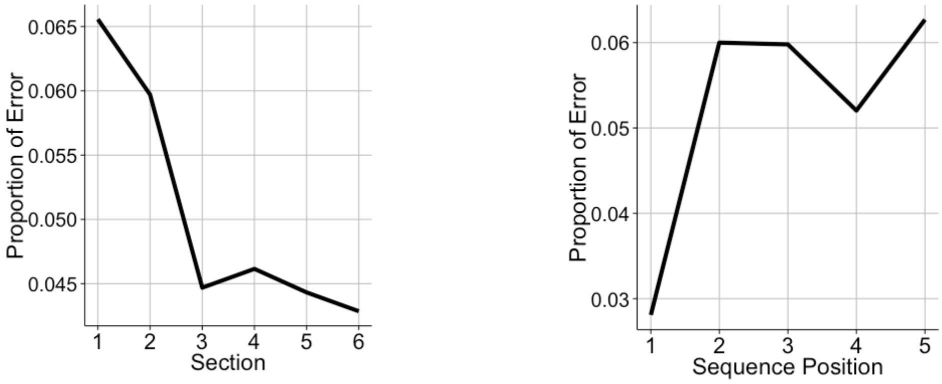
First we examined whether participants learned the grammatical knowledge in the language. Participants became faster as the experiment progressed, β= -0.01, *SE* = 0.002, $\chi^2(1)$= 14.93, $p < .001$. Also, there was a marginal interaction between section and sequence position, showing a trend towards a decrease in speed in later sequence positions, β= 0.001, *SE* = 0.001, $\chi^2(1)$= 3.33, $p = .07$.



(a) **(b)**

Figure 2.22. Average reaction time in different sections (a) and at different sequence positions (b) in comprehension (solid line) and sequence-recall task (dashed line).

Error analysis showed overall reduction in errors over the course of the study, β= -0.15, *SE* = 0.04, $\chi^2(1)$= 20.06, $p < .001$. Participants made more mistakes in later sequence positions, β= 0.17, *SE* = 0.04, $\chi^2(1)$= 15.89, $p < .001$, and this suggests that position-specific expectations were influencing their choices.

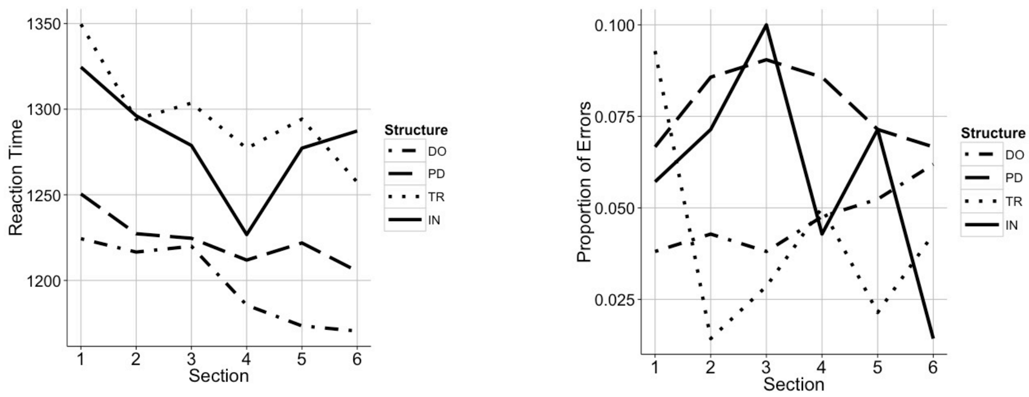


(a) **(b)**

Figure 2.23. Average reaction time in different sections (a) and at different sequence positions (b) in comprehension (solid line) and sequence-recall task (dashed line).

An analysis of only INOUN category letters in different structures was then performed. Reaction times showed that participants responded differently to INOUN letters in different structures, $\chi^2(1)$= 51.61 (3), p < .001. No other effects or interactions were observed.

Accuracy analysis showed a main effect of structure, showing that error rates were different for different rules, $\chi^2(1)$= 16.46 (3), p < .001. No other effects or interactions were observed. Thus participants learned grammatical knowledge of this novel language, which was evident in their processing speed and accuracy.



(a) **(b)**

Figure 2.24. Reaction times (a) and error rates (b) when selecting INOUN category letters in different structure sequences over the course of the study.

5.2.1. Verb bias and reaction times

To test the effects of verb bias on comprehension speed, reaction times were taken to find the post verbal letter and were predicted by section crossed with verb-structure match. The model also included random slopes for section crossed with verb-structure match. The results revealed that participants were faster when the verb occurred in its preferred structure β= -0.06, *SE* = 0.02, $\chi^2(1)$= 40.4, $p < .001$. No other main effects or interactions were found.

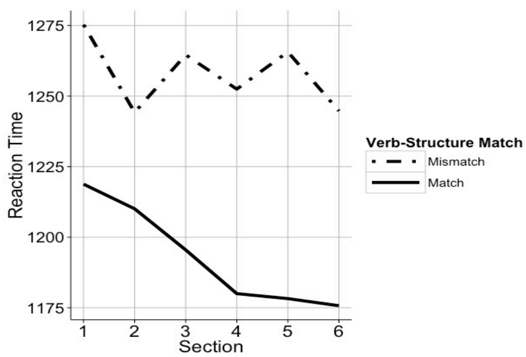


Figure 2.25. Mean reaction times at different sections in comprehension task when verb occurred in its preferred (solid line) or non-preferred (dashed line) structure.

5.2.2. Verb bias and accuracy

Error rate in comprehension was 6.6% (SD=2.5). Error analysis in comprehension revealed no main effects or interactions. Since sequence completion trials contained novel verbs, verb bias was not examined in this task.

5.2.3. Structural priming and reaction time

To test the effects of prime structure on the target sequence processing, reaction times taken to find the post verbal letter were predicted with section crossed with prime-target match. The model included the same variables in its random effects structure. The results showed that participants were faster to find the post verbal letter of the target sequence when its structure matched that of a prime sequence (β= -0.06, *SE* = 0.02, $\chi^2(1)$= 40.4, $p < .001$.

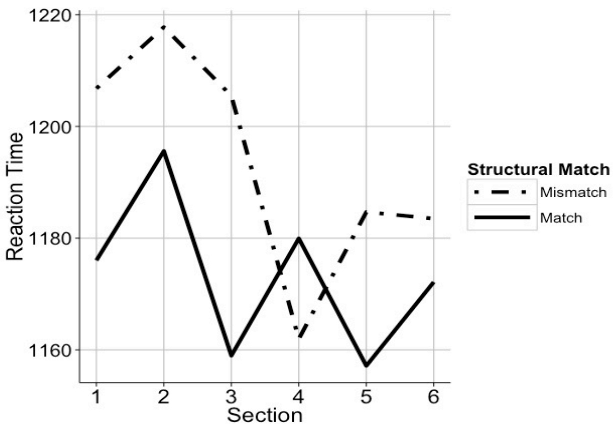


Figure 2.26. Mean reaction time at different sections in comprehension task when target sequence was preceded by the same (solid line) or different (dashed line) structure prime.

The reaction time analysis in sequence completion task revealed that participants became faster as the experiment progressed, β= -0.05, *SE* = 0.007, $\chi^2(1)$= 32.55, $p < .001$. No other effects were observed.

5.2.4. Structural priming and accuracy

Analysis of accuracy data in sequence completion task revealed no significant main effects or interactions.

5.2.5. Structural priming and structure choice

To test the effects of prime structure on participants' choices of the target sequence structure, proportion of produced PD structures was predicted by section crossed with prime structure (PD or DO, effects coded). Random slopes for the same variables were included in the model's random effects structure. The analysis revealed a two-way interaction between section and prime structure, showing that participants began producing more PD structures following PD structure primes as the experiment progressed, $\beta = 0.35$, $SE = 0.10$, $\chi^2(1) = 10$, $p < .002$.

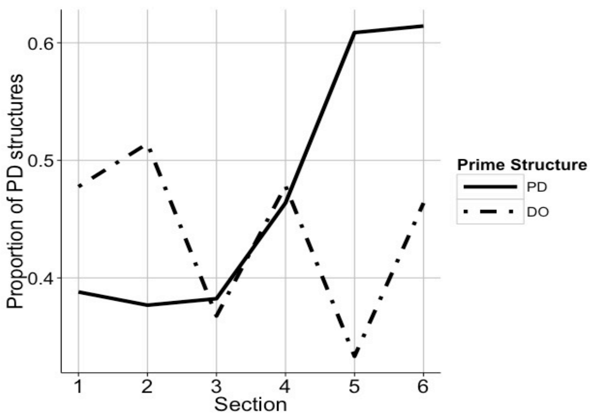


Figure 2.27. Proportion of produced PD structure sequences at different sections in sequence completion task when the prime sequence has PD (solid line) or DO (dashed line) structure

5.2.6 Priming and verb-structure match in prime sentence

To test if this priming effect was modulated by the match between sequence structure and verb bias in the prime sequence, the same analysis as above was repeated but with an addition of prime verb type (DVERBP or DVERBD, effects coded) that was fully crossed with section and structural match between prime and target (match, mismatch, effects coded) as a predictor variable.

The results showed a main effect of section, indicating that participants produced more PD structures as the experiment progressed, β= 0.11, *SE* = 0.06, $\chi^2(1)$= 4.33, *p* = .04. Just like in the previous analysis, there was an interaction between section and prime structure, showing that PD structure primes led to increasingly greater proportion of completed PD target sequences as the experiment progressed, β= 0.38, *SE* = 0.10, $\chi^2(1)$= 16.71, *p* < .001. Finally, there was a two-way interaction between prime structure and verb type, providing preliminary evidence that participants produced more PD structures when the prime had a PD structure that contained a DO biased verb (β= 0.69, *SE* = 0.33, $\chi^2(1)$= 3.89, *p* = .05).

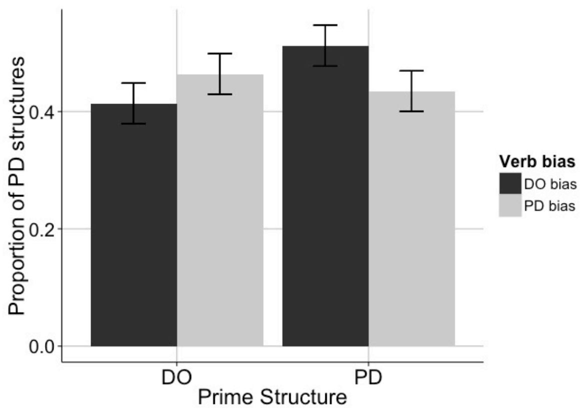


Figure 2.28. Proportion of produced PD structure sequences in sequence completion task when the prime sequence has PD or DO structure and either DO biased verb (dark grey) or PD biased verb (light grey)

5.3 Summary of Experiment 4

The section/position/INOUN analysis of the errors and reaction times provides evidence that participants learned the grammatical constraints of the language. Verb bias effects were evident in comprehension speed where participants were faster at finding the post verbal letter when verbs occurred in their preferred structure.

In this study, we found the first structural priming effects in comprehension reactions times, where participants were faster at finding the post-verbal letter of the target sequence when it was primed with the same structure sequence. The effect also translated into structural choice where over the course of the experiment people began producing more PD structure sequences following a prime sentence with a matching structure.

The central outcome of this study was the interaction between prime structure and verb bias, where the structural priming effect was stronger when the verb's structural preference mismatched the structure that it appeared in. This is consistent with the findings in natural language sentence production tasks (Bernolet & Hartsuiker, 2010; Peter, Chang, Pine, Blything & Rowland, 2015) and supports mechanisms that use prediction error to explain priming (Chang et al., 2006). This can also potentially explain why this study found an effect of priming in comprehension speed, which was not the case in the previous experiment that used a similar method. Since the magnitude of priming is associated with the degree of surprise caused by the processing of the prime structure, stronger priming, and, in turn greater learning, is expected when sequences contain verbs that are biased towards a competing structure (Fine & Jaeger, 2011; Jaeger & Snider, 2007). In Experiment 2 and 3, all prime sequences contained verbs that matched the structure that they appeared in, which could explain why the same effects were not observed in those experiments.

6. General discussion

The goal of this research was to examine whether it was possible to simulate the acquisition of English-like sentence structures in a non-linguistic statistical learning task, looking at whether the same domain-general statistical learning mechanisms could be used to explain verb bias and structural priming effects. We examined both the acquisition of English-like structures and structural adaptation effects within an SRT task. Across the four experiments, we found evidence that participants learned structural aspects of a novel artificial language, as indicated by faster and more accurate sequence production as the study progressed. Although participants showed general improvements on the task, they also showed improvements that reflect grammatical knowledge. This is most clearly seen in the way that the participants responded differently to the same category (INOUN) letters presented in different sentence positions in different structures. Since the INOUN letters were in the same position on the circle in the last 3 studies and the task involved simply moving the mouse on the target letter, there is no reason for differences in the processing of these letters unless participants were encoding structure-specific expectations for the INOUN letters at different sentence positions.

The fact that participants learned the structure of the language is also supported by the verb bias effects found across the four studies (Table 2.6). With the exception of experiment 1, in the comprehension-like tasks, participants were faster at finding the post verbal letter when the structure of the sentence was consistent with the structural preference of the verb. This shows that participants learned the probabilistic constraints of the language using the knowledge of verbs' structural preferences to guide online structure processing. Such an increase in processing speed is similar to the verb bias effects found in natural language comprehension studies (Garnsey et al., 1997, exp. 1; Kennison, 2009, exp. 1). Verb bias effects were also evident in sentence-recall task in experiment two and the sentence-completion task in experiment three. These findings mimic the verb bias effects seen in sentence-recall accuracy studies (e.g. Lombardi & Potter, 1992) and language production speed studies (Gahl & Garnsey, 2004; Jennings et al., 1997; Stallings et al., 1998) respectively. Participants learned the biases during the comprehension-task and showed the transfer of that knowledge to production-like sentence-recall and sentence-completion tasks. Thus different tasks shared the same representations of the structural preferences of the verbs. This is especially evident in experiment three, where participants always had a free choice of which structure to produce given a verb and they showed clear biases in their choices that came from the comprehension task. Similar effects are observed in production norming studies (Garnsey et al., 1997; Trueswell et al., 1993). From these studies, it is clear that verb bias is robustly learned across different tasks from statistical regularities in the linguistic input and the acquisition of this knowledge does not require specialized language-specific constraints.

An interesting question, however, is to what extent verb bias-like effects reflect the structural knowledge and to what extent they reflect simpler directional preference or motor movement biases that are based on the particular verb position. In other words, whether performance reflects the knowledge that one simply has to go to the left or to the right or the knowledge that individual letters have a role in predicting of some deeper structural knowledge about the sequences. The support that it reflects structural knowledge is provided by the interaction between verb's structural preference and the structure of the prime and its effect on structural priming which is discussed in the context of error-based learning later in the discussion.

Structural priming requires abstract structural representations and we found evidence that participants learned the grammar well enough for priming to occur (Table 2.7). Experiment two showed that participants' accuracy improved over the course of the study selecting the post verbal letter of the target sentence that was primed with the same structure sentence. Improved accuracy due to priming was also observed in the sentence-completion task in experiment three and replicated in experiment four. Across these studies, it is clear that structural priming is more sensitive to task constraints. In natural language studies, dative priming is robust in sentence completion (Pickering & Branigan, 1998), picture description (Bock 1986), but occasionally weaker in RSVP (a marginal dative priming effect in Tooley & Bock 2014, although see Potter and Lombardi, 1998, and Chang et al. 2003 for significant dative priming effects). Thus, there is variability in structural priming across different tasks and this is compatible with the results in our SRT task.

Table 2.6.

Verb bias. Bold when result is in the non-predicted direction

Task Type	Type	Effect	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Comprehension	Accuracy	V bias		**$\chi^2 = 3.38$, $p = .07$**	**$\chi^2 = 3.67$, $p < .001$**	
		Section:V bias	$\chi^2 = 3.02$, $p = .08$			
	RT	V bias	**$\chi^2 = 19.4$, $p < .001$**	$\chi^2 = 28.53$, $p < .001$	$\chi^2 = 26.28$, $p < .001$	$\chi^2 = 40.4$, $p < .001$
		Section:V bias	$\chi^2 = 3.38$, $p = .07$			
Sentence recall/ completion	Accuracy	V bias	n/a	**$\chi^2 = 5.53$, $p = .02$**		n/a
		Section:V bias	n/a			n/a
	RT	V bias	n/a	$\chi^2 = 13.9$, $p < .001$	$\chi^2 = 8.35$, $p = .004$	n/a
		Section:V bias	n/a			n/a
	PD Proportion	V type	n/a		$\chi^2 = 6.68$, $p = .01$	n/a
		Section:V type	n/a			n/a

Table 2.7.

Structural Priming. Bold type denotes effects that were in the non-predicted direction

Task Type	Type	Effect	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Comprehension	RT	Match	**$\chi^2 = 3.4$, $p = .07$**			$\chi^2 = 40.4$, $p < .001$
		Section:match				
Sentence recall/	Accuracy	Match	n/a		$\chi^2 = 24.42$, $p < .001$	
completion		Section:match	n/a	$\chi^2 = 2.89$, $p = .09$		
	RT	Match	n/a			
		Section:match	n/a			
	PD	Prime type	n/a			
	Proportion	Section: Prime type	n/a		$\chi^2 = 3.75$, $p = .05$	$\chi^2 = 10$, $p < .002$
	Bernolet	PD prime:DO verb	n/a	n/a	n/a	$\chi^2 = 3.89$, $p = .05$

From a methodological point of view, the fact that we observed language learning in such a short non-linguistic task is particularly intriguing considering that other studies tested verb bias (e.g. Wonnacott et al., 2008) or category formation (Hunt & Aslin, 2010) effects using extensive training and testing sessions spread across 5 days. This was partly achieved by removing testing sessions and measuring changes in participants' performance as they learned the language (as in Cleeremans & McClelland, 1991). This provided both an online measure for studying learning and processing together and allowed shorter training regimes. Another reason was that, unlike word form to meaning learning in Wonnacott et al's. (2008) study, or symbol learning in Hunt and Aslin (2010) study, the use of letters eliminated the need to learn the 'words' of the language before proceeding to higher order structural learning. Furthermore, providing animacy categories and highlighting the letters on the circle allowed participants to exhibit their anticipatory grammatical knowledge. These changes allowed us to use more complicated structures that closely match those in English and to examine phenomena like structural priming that have not been found in previous tasks.

The present series of studies was motivated by the idea that simple recurrent networks are able to explain both non-linguistic sequence learning in SRT studies (Cleeremans & McClelland, 1991) and language learning and processing related phenomena (Elman, 1990). These networks have been shown to acquire syntactic knowledge (Chang, 2002; Elman, 1990, 1993), learn verb biases (Juliano & Tanenhaus, 1994; Twomey et al., 2014), and exhibit structural priming (Chang et al., 2006) using the same underlying statistical learning mechanism. Thus, it should be possible to use a SRT learning task to model behaviourally a link between syntax acquisition, verb bias, and structural priming. Simple recurrent networks view learning as prediction about incoming information. The model makes a prediction about the next symbol in the sequence and uses the error between its prediction and the actual next symbol to adjust the weights in the network. These weight changes lead to the creation of abstract syntactic categories, because these categories are useful for predicting word sequences. Verbs are also useful for predicting structural continuation of sequences and thus verb biases are also learned and stored in the network (Twomey et al., 2014). Finally, if the network is kept in its learning mode during sentence processing, weight changes occurring during sentence processing will create stronger expectation of the same structure to be encountered in the future leading to structural priming (Chang et al., 2006). Together, these mechanisms predict that verb bias and structural priming should interact. When the verb bias creates an expectation of one structure but then the structure mismatches its expectation, it creates a large prediction error, leading to larger weight changes in the network and in turn greater expectation of the same structure, which leads to more priming (Bernolet & Hartsuiker, 2010; Fine & Jaeger, 2011; Jaeger & Snider, 2007). Similar effects were observed in experiment four and this suggests that structural priming in this SRT task was supported by some computational mechanism that is based on expectation and that resembles error-based learning.

This interaction between verb bias and prime structure can also shed some light on whether verb bias effects in this study are not simply based on some general preference to go to the left or to the right after particular verb letter without the need to learn deeper structural knowledge about the role of these verbs in predicting structural information. Although this effect is not very strong and only provides preliminary evidence that will need to be replicated.

Structural priming effects in this artificial language task were more robust in production-like tasks than comprehension-like tasks. We failed to find abstract structural priming effects in comprehension in the first three experiments, but succeeded in the final study. In natural language, comprehension priming studies sometimes fail to find priming without verb overlap (Tooley & Traxler, 2010; Arai et al., 2007; Tooley & Bock, 2014). And while Tooley and Bock (2014) argue that production and comprehension priming reflect similar structural mechanisms of language processing, they found some modality differences in dative sentences. Despite the use of the same materials, procedures, and participants in production and comprehension tasks, production priming was found to be robust, but priming in comprehension was found only in dative sentences with the 'to' preposition (e.g. The company rented a house to the homeless family) but not with the 'for' preposition (e.g. An inventor built a radio for his mother). Differences in natural language comprehension and production could be due to specialization in the systems that support these processes, either in evolution or in language development. However, since this non-linguistic task is too short to support the segregation of comprehension-like and production-like representations, the fact that we find that task constraints create a similar amount of variation suggests that variation in natural language might also be explained by the task constraints on production and comprehension.

6.1. Limitations

Some caution, however, should be taken when discussing the extent to which abstract structure learning in the present task is comparable to real language. While this and previous studies (e.g. Hunt & Aslin, 2010) have demonstrated abstract category learning, it is important to note that the categories in our artificial language comprised only a few elements, whereas, in real language, categories are often much larger. But it is also the case that learning a real human language can take many years and the simplicity of the present language is one reason that it can be learned in a short experimental session.

Another issue is that the present task is a visual-motor task and language is often thought to be in a separate module from those that do visual or motor processing (Fodor, 1983). It is possible then that language processing involves learning mechanisms which are completely different from those in our visual-motor task. While it is not possible to rule out this possibility, it is important to remember that many experimental techniques that are used to study language typically involve both visual and motor components. For instance, self-paced reading involves button presses in response to visual words (e.g., Ferreira & Clifton, 1986). Eye-tracking involves eye muscle movements to visual targets (e.g., Garnsey et al., 1997). Picture description involves visual scenes and mouth movements for speech (e.g., Bock, 1986). Computational models of language processing are often specific to the visual-motor features of a particular paradigm. For example, models of eye-tracking for reading (e.g., EZ reader, Reichle, Tokowicz, Liu, & Perfetti, 2011) do not model self-paced reading data and that suggests that the different motor component (e.g., eye vs. hand) might play an important role in how language is manifested in each task. Thus, when we try to explain language-processing behaviour, we are also modelling components related to the visual-motor components of these tasks and it is difficult to isolate the "pure" language component in experimental data. Our view is that it is important to see if there are language learning phenomena which cannot be learned within a visual-motor sequence learning task and the Circle task provides one way to test the claim that language has specialized learning mechanisms.

A related issue is the way that the results varied with different dependent measures. In particular, speed and accuracy seemed to trade-off against each other in the Circle task. Similar speed/accuracy trade-offs exist in language studies and it has been argued that participants adapt their processing to reflect the task demands (Lewis, Shvartsman, & Singh, 2013). In a typical comprehension study, participants see various linguistic stimuli and then answer comprehension questions, but there is no penalty for making errors. In production studies, participants produce structures and there is no penalty for producing inappropriate structures. Hence, speed should typically dominate accuracy in natural language processing tasks. The Circle task has a similar bias for speed because the language has no meaning and participants are not penalized for errors. This can help us to understand the data in experiments two and three, where RT showed faster processing of structures that matched the bias of the verb, while accuracy was lower in the same conditions. Future work needs to examine whether there are speed/accuracy trade-offs in natural language paradigms that mimic those in the Circle task.

6.2. Conclusion

Syntax has often been argued to be a special abstract type of knowledge that is not easily explained by domain-general learning mechanisms (Chomsky, 1956). However, psycholinguistic evidence has suggested that linguistic representations are constantly changing in children and adults and these adaptive processes can be modelled using statistical sequence learning mechanisms like simple recurrent networks. The present studies showed that humans have SRN-like mechanisms that can learn syntactic constraints in an artificial English-like language and exhibit verb bias and structural priming effects as they process the language. The way these effects were expressed was influenced by the response measure and the nature of the task (production, comprehension). Finally, we found evidence that priming was sensitive to surprisal and this is unique evidence in support of an error-based mechanism in syntactic processing and learning. This work supports the view that domain-general sequence learning can support language phenomena like language acquisition and adaptation.

Chapter 3. Input and age-dependent variation in second language learning: A connectionist account

1. Rationale for studies in Chapter 3

This thesis is examining the degree that linguistic adaptation and language acquisition can be supported by the same mechanisms. In contrast to the earlier studies that used different tasks to test linguistic adaptation and acquisition, Chapter 2 used the same non-linguistic serial reaction time artificial grammar-learning task to examine both phenomena. The task simplified some aspect of language acquisition by removing the learning of arbitrary form-sound lexical mappings and instead used letter symbols that people were already familiar with, that were 'produced' by moving the mouse cursor to the relevant spatially distributed symbols. Likewise, in contrast to previous studies that used grammatical rules that were very different from those in linguistic adaptation studies, the studies in Chapter 2 used a grammar that contained the structures which closely resembled the PD/DO alternation in English, the structures that have been extensively used in linguistic adaptation studies.

In the study, we found experiment-grain changes, where participants were faster to produce the sentences as they progressed through the experiment and this suggests that they acquired the grammatical regularities that supported the structures of the sentences. In addition, we found sentence-grain changes, where participants were more likely to choose target structures that matched the prime structures that they had experienced on the previous trial. Since the prime and target structures had different words, the transfer requires some abstract representations and Exp. 3 and 4 found evidence that these abstract representations were acquired in the course of the experiment. A plausible account of these results is that sentence-grain changes in learning particular symbol transitions also yield some higher-level changes related to structures. These sentence-grain changes build up over the study and that explains the experiment-grain learning of abstract PD/DO-like structures. The studies provide a link between sentence-grain linguistic adaptation and experiment-grain language acquisition in support of the LAMOLL account (Chang, Dell, & Bock, 2016; Dell & Chang, 2014).

While this work supports the idea that the learning processes in behaviour studies reflect the processes predicted by the connectionist models, it is possible that the experiment-grain learning of language differs from the real world year-grain learning of language. For instance, some studies show that length of language exposure, as measured by the number of years, does not predict the performance of older language learners (e.g. Johnson & Newport, 1989). This contradicts the critical notion of the LAMOLL account that experience continuously adjusts language representations leading to the creation of structural language knowledge. Thus it is not clear how the same mechanism that is responsible for sentence-grain and experiment-grain effects relate to year-grain effects natural language settings.

Providing this link is difficult because there is little long-term longitudinal data tracking how language representations change in children over their lifetime. Also in first-language (L1) learning, it is hard to measure the effect of years of input, because it is hard to disentangle the input from other developmental changes related to the child's age. Thus, in Chapter 3, we use data from L2 language learners, where it is easier to disassociate the effects of input from the effects of learners' age. While longitudinal L2 data is also difficult to obtain, it is possible to use data from multiple L2 learners to model how language input and age influence L2 learning in an average L2 learner. Since the Dual-path model has been shown to be able to use the same mechanisms to explain sentence-grain linguistic adaptation and experiment-grain language acquisition (Chang, Dell, & Bock), Chapter 3 examines whether the same model can also explain year-grain L2 learning without extensive changes to the mechanisms. If this is possible, then that will suggest that the same mechanisms can support sentence/experiment/year-grain learning, as suggested by the LAMOLL account. The following series of studies were published as a paper in Cognitive Science in collaboration with Franklin Chang.

2. Introduction

Linguistic input is critical for language learning. In first language (L1) acquisition, linguistic elements that occur more frequently are easier to learn (Ambridge, Kidd, Rowland, & Theakston, 2015; Bybee, 2006; Dazbrowska & Lieven, 2005; Marchman, Wulfeck, & Weismer, 1999; Phillips, 2006). However, the relationship between the input frequency and second language (L2) learning is less clear. Several studies have reported that the amount of language input—as measured, for example, by years living in L2 environment does not correlate highly with the acquisition of grammar and morphology in adult L2 learners who started learning the L2 at different ages (Andringa, 2014; DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Johnson & Newport, 1989; Lee & Schacter, 1997; McDonald, 2000; Oyama, 1978; Patkowski, 1980). Given that languages cannot be learned without linguistic input, these findings are counterintuitive and at odds with the notion that input plays an important role in L2 theories (Ellis, 2013; MacWhinney, 2008). This discrepancy in the role of input suggests that differences exist in the mechanisms that are used by L1 and L2 learners, and this study examines whether these differences can be explained in a unified way.

Input effects in L2 learning are modulated by the critical or sensitive period, the time window approximately between birth and puberty during which language learning is most effective (Knudsen, 2004; Lenneberg, 1967). This sensitive period effect is modulated by the age at which language learners begin learning the L2. As the age of acquisition (AoA) increases, the ability to learn the L2 decreases (Flege, Yeni-Komshian, and Liu, 1999; Johnson & Newport, 1989). While many of these AoA effects are found in explicit tasks, similar effects have been found in implicit tasks such as timed judgments (Ellis, 2005) and ERP studies (Weber-Fox & Neville, 1996). Similar AoA effects are found in L1 learning in deaf learners of sign language (Boudreault & Mayberry, 2006; Mayberry, 2010; Mayberry & Eichen, 1991) and international adoptees (Gardell, 1979; Gauthier & Genesee, 2011; Hyltenstam, Bylund, Abrahamsson, & Park, 2009). A wide range of social, motivational, input and biological factors have been proposed to explain this reduction in learning ability (for a balanced review, see DeKeyser & Larson-Hall, 2005). For these factors to explain the AoA effects, there needs to be a gradual accumulation of the negative impact of these factors, as the learner gets older (e.g., motivation to learn the L2 decreases for each year of age). Understanding the mechanism that could explain the gradual reduction in L1/L2 learning in such diverse circumstances is an important goal for understanding language learning. A classic study that investigated the sensitive period is that of Johnson and Newport (1989). The authors tested English morphosyntactic grammar knowledge in Korean and Chinese immigrants in the United States. They examined whether the English abilities of these L2 speakers could be predicted from the age at which they started learning English in immersion settings (3–39 years: age of acquisition, AoA), and years spent in the United States (7–30 years; length of exposure, LoE). The participants' L2 knowledge was assessed via a grammaticality judgment task, in which they indicated whether a given English sentence was grammatical (1a) or not (1b).

(1a) The farmer bought two pigs at the market

(1b) The farmer bought two pig at the market

The authors found that the performance dropped as AoA increased, showing that their ability to learn grammatical knowledge depended on the age at which they started learning the L2. However, they found no correlation between LoE and grammaticality judgment scores ($r = .16$, $p > .05$) and this has been replicated in several other studies (DeKeyser et al., 2010; DeKeyser, 2000; Lee & Schacter, 1997; McDonald, 2000; cf. Flege et al., 1999). The lack of LoE effect is an important issue, as it contradicts the assumption that language ability should increase as more input is experienced (Ellis, 2013).

One reason why LoE effect was not observed in Johnson and Newport's (1989) study could be related to the variation among different rules used in test sentences. The authors examined grammatical knowledge of 12 different morphosyntactic rules (Table 3.1). For example, sentence (1b) violated the plural rule use that required adding *–s* to the plural noun *pig*. Their data suggest that as AoA increased, the average grammatical knowledge dropped at different rates for different rules. Late learners performed worse with determiners and plural rules, whereas past tense and 3rd person singular rules seemed to be easier to master. Similar rule-specific effects have also been observed in several other studies (DeKeyser, 2000; Flege et al., 1999; Johnson, 1992; McDonald, 2000). Since their analyses collapsed the data over different rules, this within-subject variation could have obscured the effect of between-subject factors like LoE.

To understand the role that rule variation plays in sensitive period studies, we reanalyzed Flege et al. (1999) study, which was based on Johnson and Newport's (1989) original study but had a much larger sample of 240 Korean learners of English (compared to 46 participants in Johnson and Newport's study). To preview the findings, our analysis showed a significant effect of rule, which means that these learners were consistently better at judging grammaticality of some rules than others (consistent with rule differences in various L1/L2 studies; Leonard, Caselli, Bartolini, McGregor, & Sabbadini, 1992; McDonald, 2000; Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto, 2012; Rescorla & Reberts, 2002).

Table 3.1.

Examples of test items used to test the knowledge of 5 different grammar rules (ungrammatical rule use underlined)

Rule	Grammaticality	Example Test Item
Determiner	Grammatical	Tom is reading the book in the bathtub
	Ungrammatical	Tom is reading __ book in the bathtub
Plural	Grammatical	The farmer bought two pigs at the market
	Ungrammatical	The farmer bought two <u>pig</u> at the market
Particle verbs	Grammatical	The horse jumped over the fence yesterday
	Ungrammatical	The horse <u>jumped</u> the fence <u>over</u> yesterday
3rd person singular	Grammatical	Every Friday our neighbor washes her car
	Ungrammatical	Every Friday our neighbor <u>wash</u> her car

Past tense	Grammatical	Last night the old lady died in her sleep
	Ungrammatical	Last night the old lady <u>die</u> in her sleep

One explanation for the rule variation is the differences in the frequency with which those rules occur in the input. Higher frequency rules are thought to yield to better learning outcomes (Ambridge et al., 2015; Ellis, 2002; Lieven, 2010) and this predicts that L2 learners should be more accurate at judging the accuracy of higher frequency rules. Another explanation is that rules that are similar across the L1 and L2 are easier to learn than those that are different (L1-transfer/interference, Bernolet, Hartsuiker, & Pickering, 2013; Foucart & Frenck-Mestre, 2012; Hartsuiker, Pickering, & Veltkamp, 2004; Ionin & Montrul, 2010; MacWhinney, 2005; Sabourin, Stowe, & de Haan, 2006). One challenge for transfer accounts is that there is no agreement about how to best measure L1–L2 similarity and it would be difficult to augment the Flege et al. analysis with an objective measure of L1/L2 similarity. Therefore, to contrast frequency and transfer accounts, we performed a corpus study to quantify the input frequencies for some of the rules in Flege et al.'s study and used these frequencies in the reanalysis to understand the differences in L2 learners' performance with different rules. If the frequencies positively predicted performance in grammaticality judgment task, it would support frequency-based approaches. If this was not the case, then that would provide indirect evidence for alternative accounts like language transfer. Finally, we used a connectionist model of L1 language acquisition to see if we could model the findings in the reanalysis to understand how input frequency and language transfer might work in L2 language acquisition.

3 Corpus analysis

To make a grammaticality judgment, participants read a sentence and then classify it as either grammatical or ungrammatical. One way to make this decision would be to use knowledge about the transitions between words. For example, in the sentence *The farmer bought two pig at the market*, the transition between *two* and *pig* makes the sentence ungrammatical. One way to detect this ungrammatical transition would be to test if the frequency of the bigram *two pig* was below a threshold. However, since the raw bigram frequency can differ for different words (e.g., *twenty-three pigs* is a rare grammatical bigram), it can be hard to distinguish grammatical and ungrammatical transitions based on raw bigram frequency knowledge. An alternative statistic that automatically adjusts for this is forward conditional probability (CP), which is the raw frequency of the bigram divided by the frequency of the previous word, for example, CP = frequency of *twenty-three pigs*/frequency of *twenty-three*. There is a lot of evidence that CPs can explain infants' language learning behavior (Aslin, Saffran, & Newport, 1998; Gomez & Gerken, 2000), as well as experimental results in children/adults (Jurafsky, 2003; Levy, 2008; Monaghan, Chater, & Christiansen, 2005; Thompson & Newport, 2007). Critically, there is evidence suggesting that L2 learners show a similar sensitivity to forward CPs as L1 learners in an on-line task (Huang, Wible, & Ko, 2012). In this work, we explore whether forward CPs can explain the differences in rule performance in Flege et al.'s study. Our approach does not imply that people do not also extract other statistics such as backward CPs (e.g., frequency of *twenty-three* pigs divided by frequency of *pigs*) or 4-grams and used them to aid language use (Bannard & Matthews, 2008; Chang, Lieven, & Tomasello, 2008; French, Addyman, & Mareschal, 2011; Huettig & Mani, 2016). The goal of this analysis is to provide some evidence that rule differences are related to at least one input frequency-related measure. For example, Onnis and Thiessen (2013) compared English and Korean speakers using an artificial language learning task where the forward and backward probabilities between adjacent elements generated two equally probable and orthogonal perceptual parses of the elements. They found that English speakers preferred items with high backward probabilities, while Korean speakers preferred items with high forward transitional probabilities. This preference, according to the authors, arose from their experience with the language, as shown by the correspondence between predominant word order in each language with the direction preference of English and Korean speakers in the artificial language. Since Flege et al. (1999) study focuses on Korean learners, it is likely that these learners are collecting forward transitional probabilities as they acquire English. Therefore, we used these statistics in our corpus analysis.

To compute these statistics, we used child-directed speech from CHILDES online child language database (MacWhinney, 2000) and adult input from a spoken subset of the Corpus of Contemporary American (COCA; Davies, 2010). From CHILDES, we used the mothers' utterances (a total of 591,762 in 32 North American corpora (Bates, Bretherton, & Snyder, 1988; Bernstein-Ratner, 1984; Bliss, 1988; Bloom, 1970, 1973; Bohannon & Marquis, 1977; Brent & Siskind, 2001; Brown, 1973; Clark, 1978; Demetras, 1989; Feldman & Menn, 2003; Gleason, 1980; Hall, Nagy, & Linn, 1984; Higginson, 1985; Kuczaj, 1977; Morisset, Barnard, & Booth, 1995; Ninio, Snow, Pan, & Rollins, 1994; Peters, 1987; Post, 1994; Rollins, 2003; Sachs, 1983; Soderstrom, Blossom, Foygel, & Morgan, 2008; Suppes, 1974; Valian, 1991; van Houten, 1986; Warren-Leubecker, 1982). The remaining corpora were Cornell, MacWhinney, McCune, McMillan, Snow, and Tardif (MacWhinney, 2000). In these corpora, mothers are talking to children up to 8 years of age, as well as other adults/children (e.g., investigator, father, grandparents, siblings, uncles/aunts, babysitter). Conditional probabilities depend on rule frequencies. To compute these frequency counts, we created search terms that were based on the items used to test grammaticality in Flege et al.'s study. For example, determiner (DET) knowledge was tested with an ungrammatical sentence like *The boy is helping the man to build house*, which requires the knowledge that the verb *build* must be followed by a determiner *the* before using the noun *house*. Thus to judge the grammaticality of the sentence participants could use knowledge about how likely a verb is followed by a determiner. To calculate this, we extracted the frequency of verbs followed by determiners (verb-determiner) and the overall frequency of verbs (verb frequency) using the corpora tiers that were coded for syntactic categories and morphology. The DET rule CP was then calculated by dividing verb-determiner frequency by the verb frequency and this tells us out of all verb uses in this corpus, what proportion were followed by a determiner. In addition to the determiner rule, we also collected CPs for four other rules: plural (PL), particle use in phrasal verbs (PAR), third person singular verb inflection (3PS), and past tense (PST). The PL CP was calculated by dividing the number of plural nouns by the total number of nouns, which provided a measure of how likely a plural rule was to be encountered in the input compared to other noun forms. The PAR CP was thus calculated by taking the frequency of verbs followed directly by a particle and dividing it by the total number of verbs, and this probabilistic knowledge could help to identify non-adjacent particles as ungrammatical (e.g., *The man climbed the ladder up carefully*). The 3PS CP was calculated by dividing the number of verbs in 3rd person singular form by the total number of verbs, and this could help identify how likely a 3PS form was to be encountered. The PST CP was calculated by dividing the number of past tense verbs by the total number of verbs, and this provides information about how likely past tense was in general. Table 3.2 shows the implemented CLAN search terms (MacWhinney, 2000) and the corresponding raw frequency for each rule (number of utterances that matched).

Table 3.2.

Corpora search terms and raw frequency (number of matching utterances in CHILDES) for different rules

Rule	Search Term (example utterances)	Raw frequency
DET	+tMOT +t%mor +u +sdet|* (see if we can build a tower)	159107
PL	+tMOT +t%mor +u +sn|*-PL (that's what the chickens say)	40171
PAR	+tMOT +t%mor +u +sadv|* (you can sit some people down here)	133958
3PS	+tMOT +t%mor +u +sv|*-3S (the square goes in the square)	16570
PST	+tMOT +t%mor +u +sv|*-PAST (look what happened here)	6049
VERBDET	+tMOT +t%mor +u +sv|*^det|* (see if we can build a tower)	42038
VERBPAR	+tMOT +t%mor +u +sv|*^adv|* (go ahead)	28571
VERB	+tMOT +t%mor +u +sv|*	334191

	(look at that)	
NOUN	+tMOT +t%mor +u +sn\\|* (it's a chicken)	320650

Table 3.3 shows rule conditional probabilities for the same rules. It also includes rule CPs extracted from a subset of the COCA corpus to show that the results are consistent across different corpora. The correlation between rule CPs in the CHILDES and COCA corpora was high (r = .74), which means that the frequency of these five rules was similar across both children- and adult-directed speech. This correlation is due to the fact that the CPs for the DET/PL rules are higher than the 3PS/PST rules in both corpora, but the rank order within these rules is not always consistent. Since the COCA corpus was a transcription of television news programs (e.g., discussions of the Peacemaker missile system), we view this as being less typical of the input that L2 learners are generally exposed to in day-to-day settings. Since the CHILDES corpora include conversational speech between adults and other adults, as well as children up to 8 years of age, we view them as a better measure of the frequent word and structures that L2 learners are likely to use and know, and hence the following analyses used the rule CPs from the CHILDES corpora only.

Table 3.3.

Rule CP in CHILDES and COCA corpora

Rule	Formula used to calculate Rule CP	CHILDES CP	COCA CP
DET	VERBDET/VERB	0.126	0.14

PL	PL/NOUN	0.125	0.21
PAR	VERBPAR /VERB	0.085	0.13
3PS	3PS/VERB	0.05	0.08
PST	PST/VERB	0.018	0.11

This corpus analysis has provided two measures of frequency for each rule: raw frequency and CP. In the next section, we will test these different measures to see which best explains the rule differences in the Flege et al. study. If there is a significant positive effect of either frequency measure, then that would suggest that the 240 participants in that study had better knowledge of rules that were frequent in the input.

4. Flege et al. (1999) reanalysis

Flege and his colleagues investigated the knowledge of English grammar in 240 Korean immigrants living in the United States who had migrated at the ages between 1 and 23 (M = 12, SD = 5.9). At the time of testing, their average age ranged from 17 to 47 (M = 26, SD = 6). All participants had lived in the United States from 7 to 30 years (M = 14.6, SD = 4.6). Half of the participants were males or females and different AoA groups had a representative sample of participants with different LoE (Table 3.4).

The authors tested morphosyntactic knowledge for 10 rules using a grammaticality judgment test consisting of 144 sentences. The items were designed so that each grammatical sentence had an ungrammatical counterpart that violated a certain grammar rule (see Table 3.1 for examples). The participants heard a recorded sentence and were required to indicate if it was permissible in the English language. Consistent with Johnson and Newport's (1989) results, Flege et al. (1999) found that the scores for different rules varied with AoA, but their analysis involved separate ANOVA models for each rule. The novel feature of our reanalysis is to include rule-related predictors in the model to factor out rule variation from individual variation in LoE and AoA. In addition, we used logistic mixed effects models that could predict binary grammatical judgments for individual sentences while factoring out participant and test item variation. Since our goal was to examine how input variation influenced the acquisition of different L2 rules, we excluded the data from native English speaker and only used the data from the five rules (DET, PL, PAR, 3PS, PST) for which we had objective and comparable search terms. Since grammatical sentences must conform to multiple grammatical rules, we used the ungrammatical test items, because the correct rejection of these rules is more likely to relate to the rule that was used to make the sentence ungrammatical. There were eight test sentences for each rule (except for PAR which only had 6 items) and overall there were 9,120 judgments for the 38 test items over 240 participants.

Table 3.4.

Number of participants in different AoA and LoE groups

LoE groups	AoA groups			
	1-5	6-11	12-17	18-22
7-14	6	35	54	32
15-30	42	37	18	16

To replicate the earlier studies that found no effect of LoE, we first analyzed the data without including any rule-related predictors. Grammaticality judgments (grammatical = 1, ungrammatical = 0) were predicted by a logistic mixed model with AoA crossed with LoE (all predictor variables were centered) and participant and test sentences as random effects. The maximal model that converged contained AoA crossed with LoE as random slopes for test sentence (R version 3.0.2; Barr, Levy, Scheepers, & Tily, 2013). Likelihood-ratio tests were used to compare models and a chi-squared statistic for the comparison was used to compute p-values. The same approach was used for all the models in this paper. As seen in Fig. 3.1A, there was a significant effect of AoA which suggests an age-related reduction in L2 learning ability (β= -0.2, SE = 0.02, χ^2 (1) = 65.98, p < .001). There was no effect of LoE (p = .17) and no interaction between the two variables (p = .25). Thus, we find that the years of input is not a strong predictor of grammaticality judgments when the variability between rules is treated as unexplained variance.

Next, we added rule as a categorical factor (fully crossed with AoA and LoE) to see if L2 learners showed consistent patterns in their knowledge for certain rules. The maximal model that converged contained no random slopes. There was a significant negative effect of AoA (β= -0.161, SE = 0.02, χ^2 (1) = 177.51, p < .001), a positive effect of LoE (β= 0.001, SE = 0.03, χ^2(1) = 4.13, p = .042), and a negative effect of rule (χ^2(1) = 24.28, p < .001). There was a marginal interaction between AoA and LoE (β= 0.003, SE = 0, χ^2(1) = 3.08, p = .079). There was also a significant interaction between AoA and rule (χ^2(1)=61.78, p < .001). Finally, there was a three-way interaction between AoA, LoE, and rule (χ^2(1) = 13.56, p = .0088). This analysis demonstrates that participants with different AoA and LoE show consistent differences between the rules that they are tested on (e.g., judgments of past tense rule items were consistently better than judgments of determiner rule items). When this rule-related variability was factored out, then LoE showed a significant positive effect, where more years of input led to better knowledge of English grammar. Thus, the weak nature of LoE effects in previous studies could be due to the fact that earlier analyses treated rule variation as unexplained variance. The variation due to rule can be clearly seen in Fig. 3.1B, where we split AoA into early learners (<12 years) and late learners (>12 years, both 120 participants). We used 12 years because this is where a non-linearity occurs in the data (Flege et al., 1999), but we make no claim about the special role of this particular age.

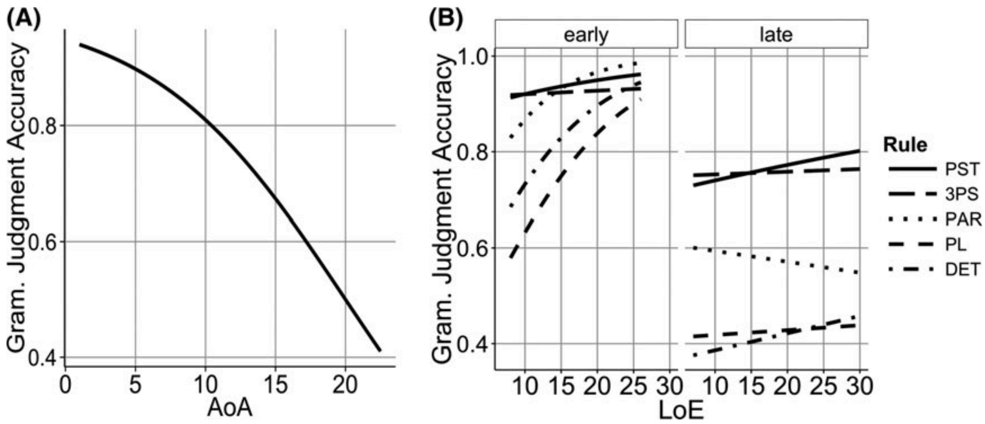


Figure 3.1. (a) Effect of age of acquisition (AoA) on grammaticality judgment scores. (b) Effect of length of exposure (LoE) on different rules in early (<12) and late (>12) AoA learner.

The above analysis suggests that there are consistent differences among the rules, but since rule is a factor, each level of rule is treated as an arbitrary category and the analysis provides no possible cause of these rule differences. One possible explanation of these rule differences is that participants rely on the knowledge of the raw frequency of the categories at the critical point in the test utterances. For example, knowing how frequently a verb is followed by a preposition can help to identify the error in the PAR rule item *The horse jumped the fence over yesterday*. To test this hypothesis, we tested a fully crossed model with categorical rule replaced by centered frequency for the adjacent categories at the critical point. The maximal model that converged contained random slope of AoA for test sentence and no slopes for participant. There was a significant negative effect of AoA (β= -0.2, SE = 0.02, $\chi^2(1) = 68.8$, $p < .001$), a marginal effect of LoE (β= 0.05, SE = 0.02, $\chi^2(1) = 3.7$, p = .055), and a negative effect of frequency (β= -0.00001, SE = 0.000003, $\chi^2(1) = 4.96$, $p < .03$). There was a marginal interaction between AoA and LoE (β= -0.006, SE = 0.004, $\chi^2(1) = 2.99$, p = .08). There was also an interaction between AoA and frequency (β= -0.0000005, SE = 0.0000002, $\chi^2(1) = 6.29$, p = .012). Finally, there was a three-way interaction between AoA, LoE, and frequency (β= -0.00000005, SE = 0.00000003, $\chi^2(1) = 3.84$, p = .05). This analysis suggests that the rule differences in judgment behavior can be explained by a frequency measure. But unlike the previous model with rule as a factor, this model found only a marginal effect of LoE. Furthermore, since raw frequency will vary with the frequency of the component categories and the size of the corpus, we will test whether forward CPs, which are less sensitive to these factors, can explain this rule variation.

The next model included forward rule CP fully crossed with AoA and LoE. Rule CPs are computed from the raw frequencies divided by the previous category and hence they can vary between 0 and 1 (regardless of the frequency of the corresponding categories or the corpus size). The maximal model that converged contained random slopes for rule CP for participants and random slopes for AoA for test sentence. There was a significant negative effect of AoA (β= -0.2, SE = 0.02, $\chi^2(1) = 68.8$, $p < .001$), a positive effect of LoE (β= 0.04, SE = 0.02, $\chi^2(1) = 4.17$, p =

.04), and a negative effect of rule CP (β= -21.3, SE = 3.76, $\chi^2(1) = 20.1$, $p < .001$). There was a significant interaction between AoA and LoE (β= -0.01, SE = 0.004, $\chi^2(1) = 4.94$, $p = .03$). There was also a marginal interaction between AoA and rule CP (β= -0.5, SE = 0.33, $\chi^2(1) = 3.37$, $p = .07$). Finally, there was a three-way interaction between AoA, LoE, and rule CP (β= -0.1, SE = 0.04, $\chi^2(1) = 5.93$, $p = .015$). This shows that as AoA increased, the weakening effect of LoE affected higher CP rules more that lower CP rules.

One puzzle in the L2 literature is that years of studying an L2 do not seem to positively predict knowledge of the L2 (DeKeyser, 2000; DeKeyser et al., 2010; Johnson & Newport, 1989; Lee & Schacter, 1997; McDonald, 2000). We replicated this finding (non-significant LoE) in our first model without any rule-related predictors. Furthermore, a model that included raw frequency did not yield a significant effect of LoE, suggesting that this predictor did not factor out rule variations sufficiently to be able to see the effects of LoE. But when rule was added as a factor or as rule CP, we found a significant positive effect of LoE, where performance improved with more linguistic exposure. In addition, while all the models exhibited a sensitive period effect (a reduction in grammatical knowledge with increased AoA), only the rule CP model exhibited a significant interaction between LoE and AoA, where late learners benefitted from the input less than early learners. We suggest that previous studies did not find positive effects of LoE or interactions of LoE with other factors because they did not fully factor out variation between rules.

In addition to clarifying the effect of AoA and LoE, these rule-related predictors in the model suggested that some rules were consistently easier than other rules, regardless of the test sentence they were in or participant differences. Both the raw frequency and rule CP models suggest that these rule differences are due to a negative relationship with frequency. This conflicts with theories of L1 and L2 learning which argue that higher frequency should lead to greater accuracy (Ambridge et al., 2015; N. C. Ellis, 2002) and this work will attempt to explain this discrepancy. To better understand this negative effect, we need to determine which measure of frequency provides the best account of the data. One way to compare these models is with R2, which is the variance explained by each model (Johnson, 2014; Nakagawa & Schielzeth, 2013). The model without rule CP explained about 21% of the variance. The model with raw frequency explained an extra 4% (R2 = .25) and the rule CP model explained about 9% more (R2 = .30). Since the rule CP model explained the most variance and uses a measure of frequency that is less dependent on word and corpus properties, we will use rule CP as our proxy for frequency in L2 learning.

The rule CP model revealed a significant three-way interaction between AoA, LoE, and CPs. This indicates that the weaker effect of LoE in later AoA learners impacted higher CP rules more than lower CP rules. Specifically, Fig. 3.1B shows that the high CP rules DET and PL have a strong positive LoE slope in early AoA learners, but the slope is smaller in late learners. However, the slopes of lower CP rules like PST and 3PS were less affected by AoA. This suggests that late AoA learners have trouble using the high frequency of higher CP rules to acquire them better.

In sum, our reanalysis of Flege et al.'s data suggested a complex set of mechanisms in L2 grammatical learning. These learners showed a sensitive period effect (negative effect of AoA). In support of frequency-based approaches (e.g., N. C. Ellis, 2002), we found that the amount of input (LoE) had a positive effect on L2 learning, but this was reduced in late learners. However, frequency-based approaches cannot explain the negative effect of rule CP, where frequent noun-based rules were associated with lower accuracy scores than less frequent verb-based rules. Since each of the 240 participants was tested on each rule, the difference in the rules cannot be easily attributed to between-participant differences in motivation, social factors, or biological factors. A likely cause of the rule differences is transfer from L1, since Korean does not have determiners and uses plural marking less than English. Support for the transfer account can be found in Ionin and Montrul (2010), who found that Korean learners of English had more trouble learning the generic interpretation of English determiners compared to matched Spanish learners, and this is presumably because Spanish speakers could use determiners in their L1 to enhance their learning of English. However, the Korean learners also learned third-person singular verbs fairly easily even though the Korean language does not mark this distinction, so it is not obvious what kind of transfer mechanism could explain the learning of this rule. One possible account of language transfer are connectionist learning mechanisms that can encode similarity structure using distributed representations (Twomey, Chang, & Ambridge, 2014). In the next section, we examine whether a connectionist model is able to explain the findings in our reanalysis.

5. A connectionist model of the acquisition of morphosyntactic rules in L2

In the present work, we developed a computational model of L2 language acquisition and sentence processing and used it to examine the results observed in our Flege et al. reanalysis. The model is based on the connectionist model of L1 learning and processing called the dual-path model (Chang, 2002). The model has several features that are relevant for its application to this dataset. First of all, the model has been shown to be able to learn abstract English grammatical constraints like those that are tested in Flege et al.'s study (Chang, Dell, & Bock, 2006). Second, the model can learn typologically different languages (Chang, Baumann, Pappert, & Fitz, 2015) and, in particular, it has been shown to be able to learn and explain various Japanese phenomena (Chang, 2009), which is a verb-final case-marked language like Korean. Finally, the model uses the linguistic input to make small changes to its morphosyntactic knowledge within a limited capacity memory and this means that the knowledge that it learns for different rules may compete with or support the learning of new rules (Fitz, Chang, & Christiansen, 2011; Twomey et al., 2014).

To simulate the environment of L2 learning at different ages, we first trained the dual-path model on Korean-like L1 input until it reached adult-like performance. The weights in the Korean model were saved after every 3,000 epochs (1,000 epochs represented one human year) and were used as the starting points for the models learning English as an L2. By varying the starting point, we simulated children who had different amounts of Korean knowledge before moving to an English-speaking environment at different ages (AoA). Since the same model weights are used to learn both languages, the model instantiates the idea that shared systems are used for both L1 and L2 languages (Hartsuiker & Pickering, 2008; Hartsuiker et al., 2004; Schoonbaert, Hartsuiker, & Pickering, 2007). This shared system assumption combined with the model's learning mechanism is consistent with evidence for transfer between L1 and L2 in various tasks (e.g., structural priming; Chang et al., 2006).

5.1. The Korean L1 and English L2 input environment for the models

Both the Korean and English languages consisted of simple intransitive, transitive, and dative structure sentences. The languages were composed of 40 words: eight animate nouns, eight inanimate nouns, six transitive, six intransitive, and six dative verbs. The Korean language included function words/morphemes (particles) that denoted case (e.g., nominative ka, accusative ul, dative ey key) and verb endings (e.g., -da). The English language contained morphemes to mark tense (-ed, -ing), third-person singular verb inflection (-ss), noun number (-z, this letter was chosen to differentiate it from third-person singular inflection), and determiners (a, an, the, this, that, two, three, many, several) with the appropriate plural counterparts. To test particle movement rules, the grammar also contained two prepositions for creating phrasal verbs (down, up).

To train the models, sentences were paired with corresponding messages. Intransitive sentences had one argument Y in the message that mapped onto the subject slot. Transitives had an agent X and a patient Y argument that mapped onto the subject and object slots, respectively. Finally, datives had an agent X, a patient Y, and a goal Z argument that mapped onto the subject, object, and indirect object slots (Table 3.5). Each argument was made up of a concept (e.g., CAT) and features that helped to structure the noun phrase (e.g., Y = CAT, THREE,DIST). There was a special argument for lexical action information (e.g., A = DANCE). In addition, the message contained event-semantics (e.g., E = PROG,YY), which had information about tense and aspect of the event. There were two possible tenses (present, PAST) with two possible aspects (simple, PROGressive). Present tense and simple aspects were considered default and had no event-semantic features. The event-semantics also contained features that encoded the number of roles that were required to describe a given event (XX, YY, ZZ). Both Korean and English languages shared the same meaning system but used different words in the lexicon to express the message. For simplicity, the Korean content word vocabulary was created by adding the letter “k” to the beginning of the English content words (the labels play no role in the model's behavior).

The language had features that captured some of the constraints in different rules in English and Korean (Table 3.6). Each noun argument in the message had a kind feature and a number feature that helped create noun phrases. The kind feature could be DEFinite, INDEFinite, PROXimate, or DISTal. The number feature could be SINGular, TWO, THREE, PLURal. All kind features were equally frequent and the singular feature was eight times more frequent than other number features. If the argument had PLUR number feature, then the noun was followed by –z (plural morpheme). PLUR nouns were preceded by the word *those* if the kind feature was DIST, the word *these* if the kind feature was PROX, the number word (e.g., *two*) if the kind feature was DEF, the word *the* if the number feature was PLUR, and nothing if the kind feature was INDEF. If the number feature was SING, then DEF mapped to the word *the*, INDEF mapped to the word *a*, PROX mapped to the word *this*, and DIST mapped to the word *that*. If the kind feature was INDEF, then the TWO number feature mapped to the word *several* and the THREE number feature mapped to the word *many* (otherwise TWO mapped to the word *two* and THREE mapped to the word *three*). If the kind feature was INDEF and number was SING and the following noun started with a vowel, then the article *a* was changed to the word *an*. If the noun was a liquid or mass noun like *sugar*, milk, *water*, or *coffee* in the plural form, then the article was omitted. In the Korean language, there were no articles except for *kthis* and *kthat*, which were signaled by the PROX and DIST features. Number features like TWO mapped to *ktwo* and THREE mapped to *kthree* in prenominal position, but there was no other plural marking. The complex nature of English noun phrase rules is one possible reason that Korean learners of English have trouble judging the grammaticality of DET and PL rules.

Table 3.5.

Examples of sentence structures used to train the model and the message that denoted the role of each constituent in the sentence.

Structure	English/Korean Sentences	Message
Intransitive	*those cat -z are dance -ing*	A=DANCE
	kthat kcat ka ksit -iss -da	Y=CAT, THREE, DIST
		E=PROG, YY
Transitive	*the cat was carrying -ing this apple*	A=CARRY
		X=CAT
	kcat ka kthis kapple ul kcarry -iss –eoss -da	Y=APPLE, PROX
		E=PAST, PROG, XX, YY
Dative	*an elk give -ss sugar to the cat*	A=GIVE
	kelk ka kcat eykey ksugar ul kgive –da	X=ELK, INDEF
		Y=SUGAR, PLUR, PROX
		Z=CAT
		E=XX, YY, ZZ

Table 3.6.

Language constraints in English and Korean

Relevant Rule	Relevant Message Features	English	Korean
DET, PL	X=DOG, DEF, SING	*the dog*	*kdog*
DET, PL	X=DOG, INDEF, SING	*a dog*	*kdog*
DET, PL	X=DOG, PROXIMATE, SING	*this dog*	*kthis kdog*
DET, PL	X=DOG, DISTAL, SING	*that dog*	*kthat kdog*
DET, PL	X=DOG, DEF, TWO	*two dog –z*	*ktwo kdog*
DET, PL	X=DOG, INDEF, TWO	*several dog –z*	*ktwo kdog*
DET, PL	X=DOG, INDEF, THREE	*many dog –z*	*kthree kdog*
DET, PL	X=DOG, PROXIMATE, TWO	*these dog –z*	*kthis kdog*
DET, PL	X=DOG, DISTAL, THREE	*those dog –z*	*kthat kdog*
DET, PL	X=DOG, DEF, PLUR	*the dog -z*	*kdog*
DET, PL	X=DOG, INDEF, PLUR	*dog –z*	*kdog*
DET, PL	X=DOG, PROXIMATE, PLUR	*these dog -z*	*kthis kdog*
DET, PL	X=DOG, DISTAL, PLUR	*those dog -z*	*kthat kdog*
PAR	A=TURNDOWN E=PAST, SIMP	*turn –ed down*	*kturndown –eoss –da*
3PS	A=TURN E=PRES, SIMP	*turn -ss*	*kturn -da*

PST	A=TURN E=PAST, SIMP	*turn -ed*	*kturn –eoss -da*
	A=TURN E=PRES, PROG	*is turn -ing*	*kturn –iss -da*
	A=TURN E=PRES, PROG	*was turn -ing*	*kturn –iss –eoss -da*

There were also rules for verb construction that depended on the event-semantic features. If the features had PROG, then the verb was followed by –ing and preceded by the word *is* if the feature PRES was active or the word *was* if the feature PAST was active. If the aspect was simple, then –ed was added after the verb for the PAST feature or –ss for the PRES feature. If the subject was plural, then the word *is* was changed to the word *are*, the word *was* was changed to the word *were*, and the –ss marking was removed. In Korean, simple PRES verbs were followed by –da, simple PAST verbs by –eoss –da, PROG PRES verbs by –iss –da, and PROG PAST verbs by –iss –eoss -da. In English, there were several phrasal verbs. There were intransitive verbs *give-up* and *show-up* that combined dative verbs *give* and *show* with the prepositions *up*. There were two transitive verbs *turn-down* and *break-down* that combined intransitive verbs *turn* and *break* with the preposition *down*. In Korean, these phrasal verbs were treated as separate verb forms. Therefore, the Korean model will have to learn that in English, verbs like *turn* can have two forms with different syntactic constraints and this should complicate the learning of the PAR rule. Although English and Korean have different rules for verbs, they are less different from each other in this respect.

The grammar was created to match the order in which the five rules occurred in the corpus analysis in terms of their CPs (Table 3.7). The CPs for these rules in the model's training set were extracted using the same formula as in the corpus analysis.

Since the language was a simplified version of English, the model input CP values only match the relative order of CPs in the human data (correlation between the two is .95). To train the models, 10 randomly generated training sets of 20,000 message-sentence pairs were created for each age of L2 acquisition. This created 10 model subjects for each different AoA group. The message was excluded from 25% of the training pairs to increase the syntactic nature of the learned representations.

5.2. Dual-path architecture

The dual-path architecture is a connectionist architecture that can learn abstract rule-like syntactic representations that interact with messages in sentence production (Chang, 2002). It has two pathways; sequencing pathway for learning sentence structure (lower half of Fig. 3.2) and meaning pathway for learning word to role mappings (upper half of Fig. 3.2). To adapt the model for L2 learning, the input and output layers have word units for the words in both English and Korean languages. Otherwise, the other features of the model are similar to the previous L1 versions of the dual-path model. The sequencing pathway is based on a simple recurrent network (SRN) architecture (Elman, 1993). The network attempts to predict the next word in a sequence from the previously heard word. The previous word is an activation pattern in the Previous Word (Input) layer. Activation spreads from the Previous Word layer to the Hidden layer via a CCompress layer and then from the Hidden layer to the Produced Word layer via another Compress layer. The function of the two compress layers is to force the model to form grammatical categories instead of learning individual word-to-word mappings (Elman, 1993). The Hidden layer learns and stores representations (activation patterns) that map between the categories of the previous word and the next word and it also receives input from a Context layer that holds a copy of the Hidden layer's activation at the previous time step (dotted arrows in Fig. 3.2). This allows the model to learn longer distance dependencies between elements (Christiansen & Chater, 1999b).

Table 3.7.

Rule CPs in English corpora and in the grammar of the model

Rule	Corpora rule CP	Model rule CP
DET	0.126	0.47
PL	0.125	0.4
PAR	0.085	0.22
3PS	0.05	0.16
PST	0.018	0.11

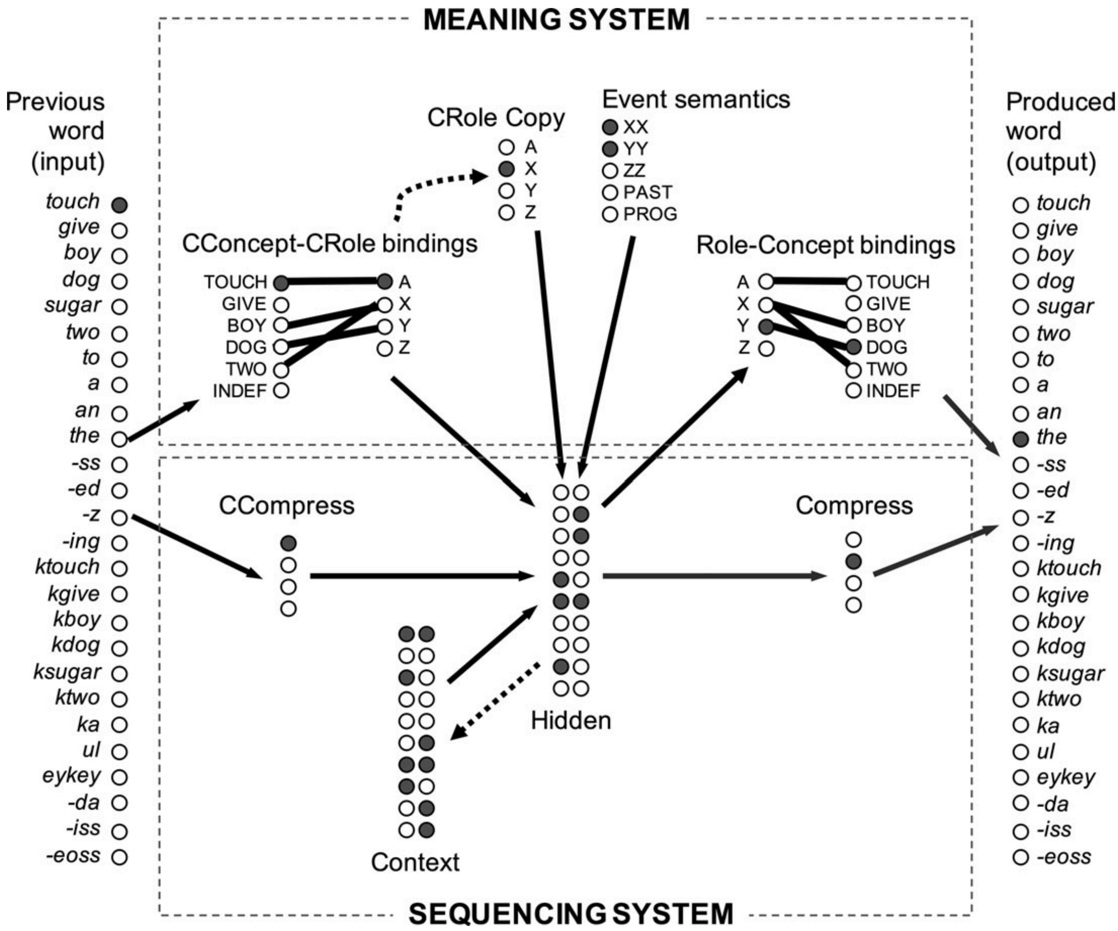


Figure 3.2. Dual-path architecture. Black/gray arrows represent connections that have to be learned via back-propagation of error. Thick lines represent fast-changing message weights. Dotted arrows show copy links.

The model learns through back-propagation of error (Rumelhart, Hinton, & Williams, 1986). At the beginning of the training, the weights are initialized randomly with a range of 0.5. First, activation spreads through the network and generates a prediction about the next word in a sentence. The mismatch between the predicted Produced Word activations and the target is called error, and it is used to make small changes in the connection weights that generated the prediction. This error signal is then propagated back through the network adjusting the connection weights between all layers so that the predicted output better matches the target. Using this mechanism, the model learns weights that encode the structure of the language (all solid arrows in Fig. 3.2).

The sequencing system interacts with the message information in the meaning system. The message is instantiated in weights between a set of Role units and the Concept layer (Role-Concept bindings). When the message contains Y = DOG, the Y role unit is linked to the concept DOG with a weight of 6 (thick black lines in Fig. 3.2). Since the Concept layer is linked to the Produced Word layer, the model can learn to activate a particular word when the appropriate concept is activated (concept DOG would activate *kdog* in Korean and *dog* in English). To allow the sequencing system to know which roles are present in the message, the Event Semantics layer has units that signal the number of roles. For example, if this layer had XX and YY units activated, that would signal to the sequencing system that it should activate the agent X Role unit after the first determiner (since English agents tend to occur early in sentences). In contrast, the Korean model would learn to activate the agent X role in sentence-initial position and would also learn to activate the subject particle *ka* afterward to mark its role. In addition, the meaning system has a comprehension message, which tells the model the role of the previous word in the sentence, which helps the model produce structural alternations (e.g., active/passive). This system maps the Previous Word layer to the CConcept layer, which is linked to the CRole layer with a reverse copy of the Role-Concept links (thick black lines on left side of Fig. 3.2). There is also a CRole Copy layer that helps the model keep track of the roles that have been processed.

In the present work, we apply the dual-path model to explain L2 behavioral data in the Korean L2 English learners in the Flege et al. (1999) study. In the present work, we train models using the Korean language as an L1 and then expose them to

English as an L2. Consistent with the claim that L1 and L2 involved the same learning mechanism, but differ in the nature and timing of the input, we have kept the L2 version of the dual-path model as similar as possible in its architecture and parameters to L1 English versions of the model (e.g., Twomey et al., 2014).

5.3. Evaluating the model's English grammatical knowledge

To gauge the overall learning of the language at different AoAs in the 10 models, we assessed the word prediction accuracy every 3,000 epochs using 200 randomly generated test sentences. To see how successfully the model learned the grammatical constraints in the rules in the Flege et al. study, we also examined its ability to distinguish grammatical and ungrammatical versions of the five rules in our reanalysis (DET, PL, PAR, 3PS, PST). Each test item had a matched grammatical and ungrammatical version (Table 3.8), and there were 100 items for each of the five rules. To test the model's knowledge of each rule, the sum of squares prediction error (the difference between the actual activation and the target activation for the word layer) for the target word at the part of the sentence where the grammatical and ungrammatical sentences differed was computed for both versions. For example, to test DET rule in the sentence *a boy touch –ed the apple*, the error of predicting the article *the* was compared to the error of predicting the word *apple* when the article was omitted as in *a boy touch –ed apple*. For each rule, the average sum of squares error (SSE) was calculated for both the grammatical and ungrammatical items. Then a rule proportion measure was computed by dividing the average SSE of ungrammatical sentences by the sum of the average SSEs for both grammatical and ungrammatical sentences. Since error levels should be larger for ungrammatical sentences than grammatical sentences, higher rule proportion scores express better rule knowledge. If the model has not developed strong expectations about whether the verbs tend to be followed by determiners or not, then SSEs for both should be similar and rule proportion should be close to 0.5. Rule proportion in the simulations approximated

Table 3.8.

Grammatical and ungrammatical sentences used to test models' performance with different rules

Rule	Error Type	Example
DET	Grammatical	*A boy touch –ed the apple*
	Determiner omission	*A boy touch –ed __ apple*
PL	Grammatical	*Two boy –z touch -ed the apple*
	-z morpheme omission	*Two boy __ touch -ed the apple*
PAR	Grammatical	*A boy break -ss down the apple*
	Particle omission	*A boy break –ss __ the apple*
3PS	Grammatical	*A boy touch –ss the apple*
	-ss morpheme omission	*A boy touch __ the apple*
PST	Grammatical	*A boy touch –ed the apple*
	-ed morpheme omission	*A boy touch __ the apple*

the grammatical judgment accuracy measure in the Flege et al.'s study and our goal is to see if the model shows similar results to those observed in the reanalysis of their data. It is known that in ERP studies (e.g., Weber-Fox & Neville, 1996), the brains of L2 learners generate mismatch signals and this means that there is evidence that implicit prediction error signals like SSE are generated in their brains and could be used to make grammaticality judgments. However, since L2 tasks vary in their dependence on implicit and explicit knowledge (R. Ellis, 2004, 2005, 2006), different tasks might have different assumptions about the way that implicit signals like SSE are used to make behavioral choices.

6. Model simulations

We present several different simulations that attempt to approximate the L2 results in the Flege et al.'s reanalysis. Our first simulation tested whether the model's activation function could create the age-dependent sensitive period. The second simulation manipulated the sensitive period by reducing the model's learning rate after puberty. The third simulation introduced different learning rates for the lexical and syntactic parts of the model. Finally, the fourth simulation implemented a model that received both English and Korean input to mimic the learning environment of many L2 learners.

6.1. Simulation 1: Activation function-based sensitive period effects

The activation function that is typically used in back-propagation has been argued to create sensitive period effects (Elman, 1993; A. W. Ellis & Lambon Ralph, 2000; Marchman, 1993; Mermillod, Bonin, Meot, Ferrand, & Paindavoine, 2012; Munakata & McClelland, 2003; Zevin & Seidenberg, 2002). In these models, activation is spread forward in the network and the net input for a unit is the weighted sum of input activations. The net activation is passed through a logistic/sigmoid activation function to create the output activation. When the weighted sum input is 0, the logistic output activation will be 0.5. On the backward pass, the output activation is compared to the target to compute the error and this error is back-propagated through the network to change the weights. The first step of this back-propagation involves the computation of the derivative of the activation function. For the logistic activation function, the derivative is highest when the output activation is near 0.5 (derivative = o (1o) when o is the output activation). The derivative of the activation function modulates the effect of error so that the same amount of error will have a larger effect on the weights when the weighted sum input is close to 0. When the weights are small, the weighted sum input to a unit will be small and the large derivative will allow relatively large weight changes. Typically weights in these models are initialized to small values early on and hence these models should be more sensitive to input early in development compared to later in the development. Knowledge learned early in L2 learning can, therefore, become entrenched and can inhibit later L2 learning (e.g., N. C. Ellis, 2013; A. W. Ellis & Lambon Ralph, 2000; Monner, Vatz, Morini, Hwang, & DeKeyser, 2013).

In previous versions of the dual-path model (Chang, 2002), the output layer used a soft-max activation function, which creates a winner-take-all bias, so that the model prefers to select only one word. To test whether the logistic activation function can create a human-like L2 sensitive period, the first simulation used this activation function for the output layer and a constant learning rate throughout the training. To aid the comparisons with the human data, the model's age was represented as the number of training trials divided by 1,000 (e.g., 1 model year refers to 1,000 training trials or epochs). We applied a learning rate of 0.1 since this level allowed the model to learn Korean to an adult level within five model years.

To examine the AoA effects, we looked at the overall word accuracy of the Korean models that started learning English at different AoAs. Fig. 3.3A shows the percentage of correctly predicted words in the Korean (gray line) and English (black lines) models that started learning the L2 at different ages. Later AoA models appeared to learn English slower but reached similar accuracy levels after 20 model years. To explore the model's grammatical knowledge with different rules over development, a mixed effect model was used to predict rule proportion scores with AoA, LoE, and rule CP fully crossed (Fig. 3.3C). All simulations contained model subject as a random intercept with random slopes for LoE crossed with Rule CP. The analysis revealed a negative effect of AoA (Fig. 3.3B), confirming that later AoA models performed worse than early AoA models (β = -0.01, SE = 0.001, $\chi^2(1) = 65.8$, $p < .001$). LoE effect showed that longer exposure to language resulted in better overall scores (β = 0.02, SE = 0.001, $\chi^2(1) = 73$, $p < .001$). There was a positive main effect of rule CP showing that the models performed better with the higher probability rules (β = 0.14, SE = 0.01, $\chi^2(1) = 73.2$, $p < .001$). There was a two-way interaction between LoE and rule CP, where higher probability rule benefited more from increasing LoE (β = 0.008, SE = 0.002, $\chi^2(1) = 16.6$, $p < .001$). Finally, a three-way interaction between AoA, LoE, and rule CP showed that this effect became stronger as AoA increased (β = 0.001, SE = 0.0003, $\chi^2(1) = 4.07$, $p = .04$).

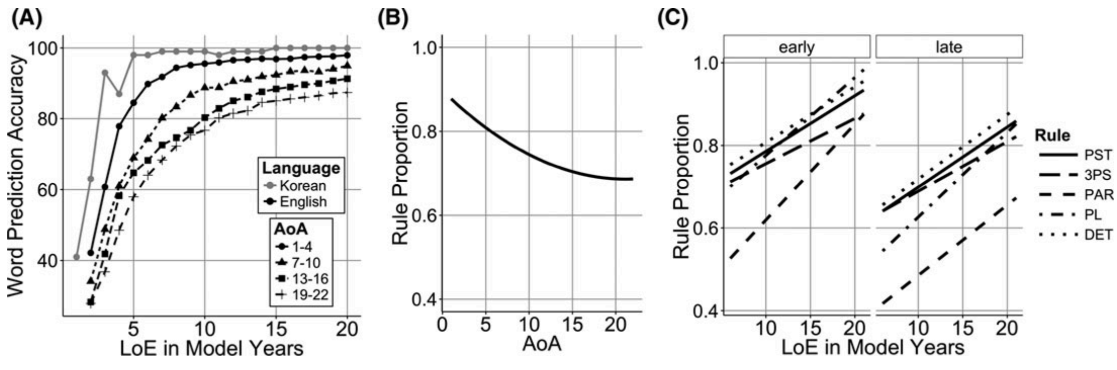


Figure 3.3. Simulation 1 model. (A) Word prediction accuracy of the Korean model (gray line) and English models that started learning English at different age of acquisition (AoA) (black lines). (B) Model rule proportion accuracy by AoA; (C) Model rule proportion by AoA, length of exposure (LoE), and Rule.

In sum, Simulation 1 showed a negative effect of AoA and this is consistent with connectionist models where the logistic function creates an age-dependent reduction in learning ability (A. W. Ellis & Lambon Ralph, 2000; Zevin & Seidenberg, 2002). However, the results of this model are different from those in Flege et al.'s (1999) data in several important ways (compare Fig. 3.1A vs. Fig. 3.3B). The sensitive period created by the logistic function is smaller than the one in human learners. Connectionist models learn from the input and therefore there is a large LoE effect in the model. Late AoA human learners in Flege et al.'s data also showed a lower sensitivity to LoE (Fig. 3.1B), but the present model shows no interaction between LoE and AoA (Fig. 3.3C). Furthermore, the human results showed a negative effect of rule CP, whereas the present model shows a positive effect. Finally, there is evidence that the sensitive period limits ultimate language attainment even with extensive input (DeKeyser & Larson-Hall, 2005), but the present model is able to catch up with early learners and hence does not match this aspect of human learning. For example, one of the participants in Flege et al. study scored only 58% judging the grammaticality of PL rule use even after 25 years of English input (model is closer to 90% at 20 model years). So while the logistic function can create age-dependent changes in learning, it does not capture the full behavior of L2 learners.

6.2. Simulation 2: Stretched Z learning rate function for the sensitive period

Simulation 1 showed that activation function was not sufficient to create a human-like sensitive period. To make the effects stronger, we directly changed the model's learning rate as it aged. There is evidence that the sensitive period has a stretched Z function (Birdsong, 2005; Flege et al., 1999; Granena & Long, 2013; Johnson &

Newport, 1989; Mayberry & Eichen, 1991), where performance is high initially, but then declines gradually and is followed by a period of slower learning. These developmental changes were incorporated into the model by keeping the learning rate high (0.1) until model year 10, after which, the learning rate dropped to 0.025 over the following 6 model years (Fig. 3.4). With this learning rate function, later learners will have a lower learning rate in development and that might keep them from changing their Korean representations to the extent that would allow them to predict English sentences with high accuracy. Also, since the previous L1 work with the dual-path model used the soft-max function on the output layer (Chang, 2002), the following simulations will use that activation function to increase the similarity between the model's account of L1 and L2 learning. Fig. 3.5A shows the percentage of correctly predicted words in Korean (gray line) and English (black lines) models that started learning L2 at a different age. While all models reached high scores with enough training, the speed with which they achieved it was slower in later AoA models. Statistical analysis confirmed that there was a significant negative effect of AoA (Fig. 3.5B), indicating that later AoA models had greater difficulty in distinguishing grammaticality (β= -0.02, SE = 0.001, $\chi^2(1) = 94.4$, $p < .001$). There was a positive effect of LoE (β= 0.01, SE = 0.001, $\chi^2(1) = 101$, $p < .001$), which showed that language exposure increased the models' accuracy, and a positive effect of rule CP (β= 0.07, SE = 0.007, $\chi^2(1) = 51.6$, $p < .001$), which demonstrated that they performed better with higher CP rules (Fig. 3.5C). There was a positive two-way interaction between AoA and LoE, showing that later AoA models benefited from increasing LoE more that early AoA models (β= 0.0005, SE = 0.0001, $\chi^2(1) = 11.6$, $p < .001$). There was also a positive interaction between LoE and rule CP, showing higher CP rules were more sensitive to increasing LoE than lower CP rules (β= 0.005, SE =

0.0001, $\chi^2(1) = 20.3$, $p < .001$). Finally, a three-way interaction between AoA, LoE, and rule CP showed that this effect became stronger as AoA increased (β= 0.0005, SE = 0.0001, $\chi^2(1) = 11.9$, $p < .001$). The reduction in the learning rate created a stronger sensitive period effect that resembles the human data more closely (compare Fig. 3.1A and 3.5B). However, like Simulation 1, the late learning models acquired the language to near-native levels (Fig. 3.5A) and the effects of rule CP and the interaction between LoE and AoA were in the opposite direction to the corresponding effects in the human data.

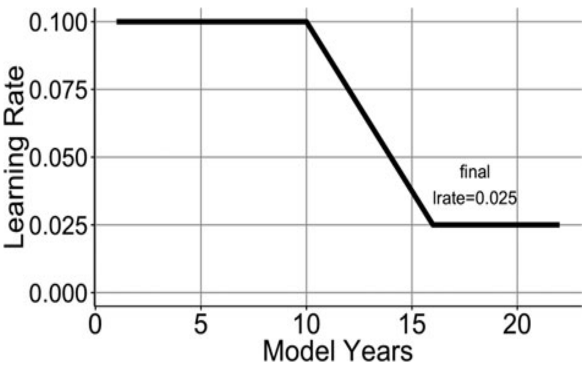


Figure 3.4. Learning rate as a function of model years.

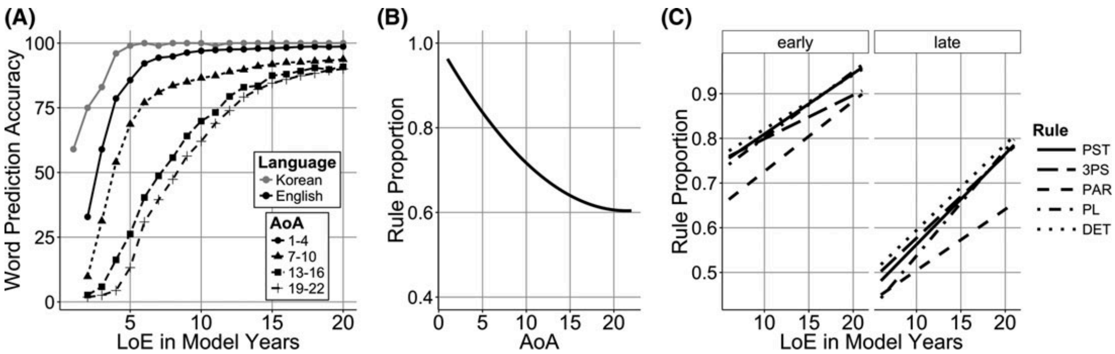


Figure 3.5. Simulation 2 model. (A) Word prediction accuracy of the Korean model (gray line) and English models that started learning English at different age of acquisition (AoA) (black lines). (B) Model rule proportion accuracy by AoA. (C) Model rule proportion by AoA, length of exposure (LoE), and Rule.

6.3. Simulation 3: Lexical and syntactic learning rates

The remaining challenge is to provide an explicit mechanism that could explain what environment creates a negative effect of rule CP. Earlier studies tried to explain rule variability in terms of intrinsic difficulty in test sentences (Bialystok & Miller, 1999), rule salience (DeKeyser, 2000), or universal factors in learnability (Johnson & Newport, 1989). However, these explanations are more speculative in nature and are difficult to confirm or falsify using statistical analysis or experimental techniques. A notable exception is the original study of Flege et al. (1999) who suggested that rules differ in terms of their reliance on lexical knowledge and more abstract rule-based knowledge. Cognitive and neurobiological explanations of the sensitive period often focus on differences between lexical and syntactic learning (Paradis, 2004; Ullman, 2015). The present study provides an opportunity to test this idea in an explicit model. The distinction between lexical and syntactic learning is supported by the studies of feral children like Genie, who started learning her first language at 13 and was able to learn new words faster than other children in the same MLU stage of development, but never fully mastered English grammatical knowledge (Curtiss, Fromkin, Krashen, Rigler, & Rigler, 1974; Curtiss, Fromkin, Rigler, Rigler, & Krashen, 1975; Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974). In addition, Singleton and Lengyel (1995) have argued that there is no sensitive period for vocabulary learning in either L1 or L2 language and in some cases, L2 learners outperform native learners in word learning tasks (Kaushanskaya & Marian, 2009). There is also evidence that late learners show N400 signatures for newly learned L2 words even after only 14 h of instruction (McLaughlin, Osterhout, & Kim, 2004). Weber-Fox and Neville (1996) found reduced syntactic P600 effects in late learners (AoA > 11) for phrase structure, but lexical N400 effects were present for both early and late learner when a word appeared in a position that was not expected in terms of meaning. These studies suggest that AoA has a greater negative impact on syntactic learning than lexical learning.

To examine this hypothesis in the model, we incorporated separate learning rates and varied them independently for the lexical and syntactic learning weights in the model. The lexical learning system included the connections between Concept and Produced Word layers and the connections between Hidden, Compress, and Produced Word layers (gray arrows in Fig. 3.2). These parts of the model were responsible for selecting the right output word, whereas the remaining parts of the model were involved in learning structural regularities (black arrows in Fig. 3.2). The syntactic learning rate remained fixed at 0.1 for the first 10 model years and then was reduced to 0 across the following 6 years. The learning rate in the lexical learning part of the system remained fixed at 0.1 throughout training (Fig. 3.6).

The focus on the distinct properties of the lexical and syntactic systems is similar to Ullman's (2001) declarative/procedural theory. In his theory, syntactic rule learning depends on implicit procedural learning and this is in agreement with our model, which only implements implicit statistical learning (Chang, Janciauskas, & Fitz, 2012). However, Ullman's theory argues that lexical learning involves declarative systems. In our model, long-term lexical knowledge is also learned through procedural learning. The fact that procedural learning is involved in lexical learning is supported by studies showing that word-based repetition priming is present in anterograde amnesia patients, even though their declarative learning systems are damaged (Gordon, 1988; Mayes & Gooding, 1989; Schacter & Graf, 1986). This type of priming has been argued to reflect implicit learning processes (Oppenheim, Dell, & Schwartz, 2010). However, the higher learning rate for lexical learning in the present simulation could help to support the fast learning of arbitrary associations and this is one of the features of the declarative memory. Thus, while this simulation has similar assumptions to Ullman's account, the model does not fully implement the declarative components of his account.

The learning rate changes in the structure learning system created a clear sensitive period effect, where later AoA models performed noticeably worse than early AoA models. However, the later AoA models were still able to use the lexical learning system to support their English grammatical knowledge and their accuracy levels approached 65% (Fig. 3.7A). Analysis of rule learning revealed that there was a significant negative effect of AoA (Fig. 3.7B, β= -0.02, SE = 0.001, $\chi^2(1) = 35.8$, $p < .001$), a positive effect of LoE (β= 0.004, SE = 0.001, $\chi^2(1) = 131$, $p < .001$) and a marginal negative effect of rule CP (β= -0.09, SE = 0.01, $\chi^2(1) = 3.1$, $p = .08$). There were also three negative interactions between AoA and LoE (β= -0.0006, SE = 0.0001, $\chi^2(1) = 4.87$, $p = .03$), AoA and rule CP (β= -0.02, SE = 0.001, $\chi^2(1) = 71.1$, $p < .001$), and LoE and rule CP (β= -0.007, SE = 0.001, $\chi^2(1) = 23.5$, $p < .001$). Finally, there was a three-way interaction between AoA, LoE, and rule CP (Fig. 3.7C), showing that with increasing AoA, higher CP rules benefited from increasing LoE less than lower CP rules (β= -0.001, SE = 0.0001, $\chi^2(1) = 37.9$, $p < .001$).

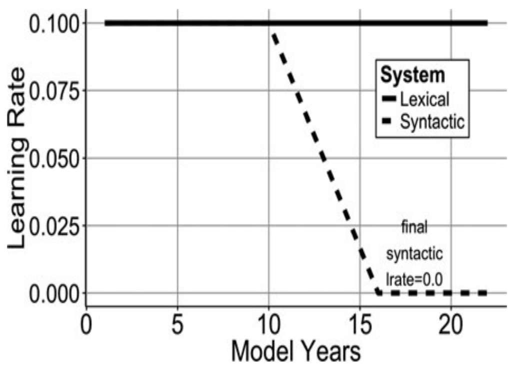


Figure 3.6. Learning rate as a function model's age in years for lexical and syntactic systems.

Separating lexical and syntactic learning parts of the system successfully captures the effects observed in the Flege et al. data. Importantly, it showed that the LoE effect was weaker in later AoA models (Fig. 3.5C). Also, the direction of the rule CP effect flipped from positive to negative. While the main effect of CP was marginal, its interaction with AoA and especially the three-way interaction between AoA, LoE, and rule CP matched the human results showing that with increasing AoA, higher CP rules benefitted from increasing LoE less than lower CP rules.

After 16 years, the model's syntactic learning rate goes to zero and therefore the late learning models are learning to predict English words using Korean syntactic knowledge. Fig. 3.7A shows that 19–22 learners do acquire the ability to correctly predict English words with an accuracy of around 70%. This relates to ERP evidence showing that late L2 learners exhibit similar syntactic P600 effects as native L1 speakers in some conditions (Foucart & Frenck-Mestre, 2011; Sabourin et al., 2006). These effects are sometimes used to argue against critical period effects since late learners are exhibiting similar patterns to native speakers. However, even though the late learning models do not have native-like L2 syntactic representations, their L1 representations are sufficient to create differences across L2 rules. This is especially the case when behavior across the whole network/brain is averaged into a single measure like Rule Proportion/ERPs, where it can appear as if human/model learners are processing L2 sentences in a native-like manner.

In this and the previous simulations, the models stopped receiving Korean language input once English was introduced as an L2. Although the complete suspension of L1 input is rare, there are many L2-dominant bilinguals (Flege, Mackay, & Piske, 2002), particularly those with early AoA with long LoE in strongly monolingual environments who would be well characterized by this model.

Furthermore, there are two populations which are similar to these models in that they show AoA effects even though they mainly receive input from one language: international adoptees and deaf learners of sign language. International adoptees are adopted into a new culture and exclusively get input from one language. Several studies have found that, while these learners have similar motivation and input to native learners, they acquire the language to a lower level than the equivalent native learners and language proficiency is negatively related to the age of adoption (Gardell, 1979; Gauthier & Genesee, 2011; Hyltenstam et al., 2009). Deaf learners of sign languages also show AoA effects, even though sign language is their L1 and they are highly motivated (Boudreault & Mayberry, 2006; Mayberry, 2010; Mayberry & Eichen, 1991). These AoA effects support DeKeyser and Larson-Hall (2009, p. 88) claim that "AoA keeps playing a large role when social and environmental variables are removed" and this suggests that some biological changes in learning ability may be involved in creating the sensitive period. Although the sensitive period is evident even when learning a single language, it is the case that most L2 learners continue to use the L1 after they start to receive L2 input and we examine whether this has an effect in simulation 4.

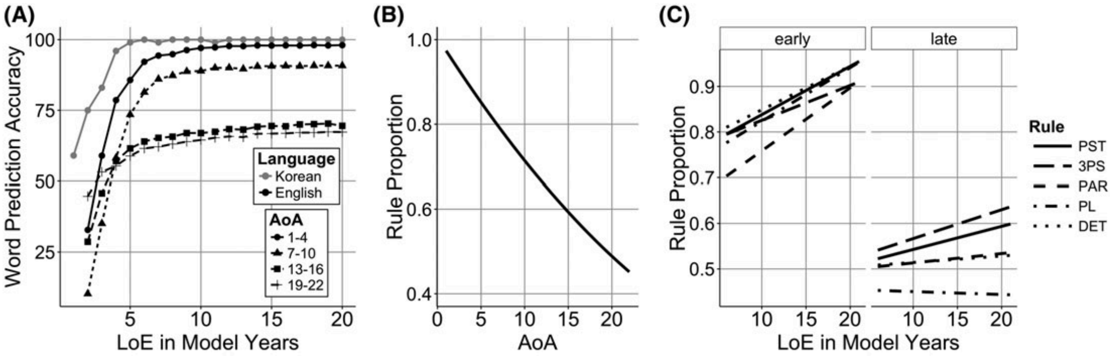


Figure 3.7. Simulation 3 model. (A) Word prediction accuracy of the Korean model (gray line) and English models that started learning English at different age of acquisition (AoA) (black lines). (B) Model rule proportion accuracy by AoA. (C) Model rule proportion by AoA, length of exposure (LoE), and Rule.

6.4. Simulation 4: Korean and English input in L2 learning

Our final simulation examines whether the results of the previous analyses generalize to an environment where the models receive both English and Korean input. Initially, the model learned Korean as an L1 and then it was given half-English and half-Korean input interleaved in a random order (akin to balanced bilinguals). To signal the target language, an additional language feature was added to the event semantics, which told the model which language it was producing. The syntactic and lexical learning rate parameters, as well as other aspects of the simulation were identical to Simulation 3.

As in Simulation 3, late learning models did not achieve native-like language accuracy (Fig. 3.8A). There was a negative main effect of AoA effect (Fig. 3.8B, β= -0.02, SE = 0.007, $\chi^2(1) = 27.5$, p < .001), a positive effect of LoE (β= 0.01, SE = 0.0004, $\chi^2(1) = 137.6$, p < .001), and a negative effect of rule CP (β= -0.08, SE = 0.007, $\chi^2(1) = 45.8$, p < .001). There was a negative interaction between AoA by LoE (β= -0.001, SE = 0.0001, $\chi^2(1) = 37.3$, p < .001), and a negative interaction between

AoA and CP (β= -0.02, SE = 0.001, $\chi^2(1) = 28.6$, $p < .001$). There was also a marginal interaction between LoE by rule CP (β= -0.002, SE = 0.001, $\chi^2(1) = 3.24$, $p = .007$). Finally, there was a three-way interaction between AoA, LoE, and rule CP (Fig. 3.8C, β= -0.001, SE = 0.0001, $\chi^2(1) = 82.8$, $p < .001$).

To better understand how bilingual input affected learning, we also examined the model's code-switching behavior (e.g., producing Korean words in English sentences) in both simulations. Fig. 3.9 shows the proportion of Korean words produced by the models that received English-only L2 training (Simulation 3) or English and Korean L2 training (Simulation 4). Late AoA models in Simulation 4 continued using many Korean words in English sentences even after a substantial number of years of English input. These results approximate the results of studies which have found that code-switching rate was higher (14%) in late learners than in early learners (6%; Sheng, Bedore, Peña, & Fiestas, 2013). Code-switching is very context dependent and this model does not fully capture all the factors that influence code-switching. For example, Moore (2013) found that English-learning Japanese speakers often switched to their L1 while preparing for an English presentation and the percentage of L1 could vary greatly within the same speaker depending on the proficiency of the interlocutor. Although AoA information was not provided for the learners in this study, there were some participants who used their L1 approximately 88% of the time, which approximates the high levels in late learners in Simulation 4.

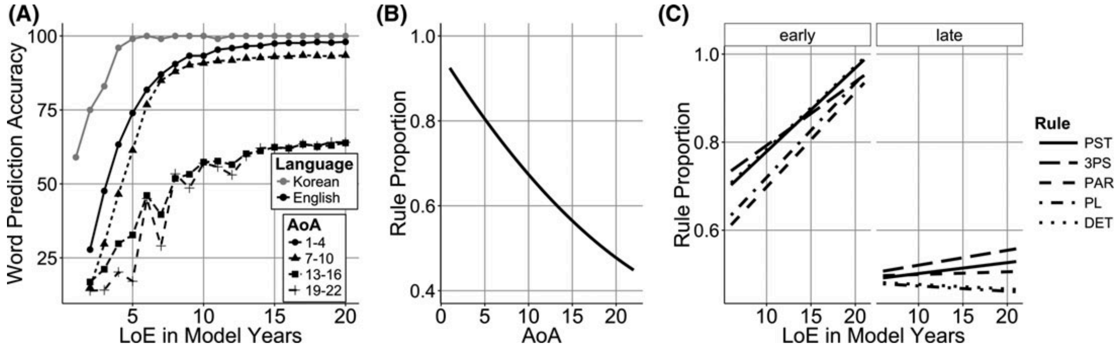


Figure 3.8. Simulation 4 model. (A) Word prediction accuracy of the Korean model (gray line) and English models that started learning English at different age of acquisition (AoA) (black lines) (B) Model rule proportion accuracy by AoA. (C) Model rule proportion by AoA, length of exposure (LoE), and Rule.

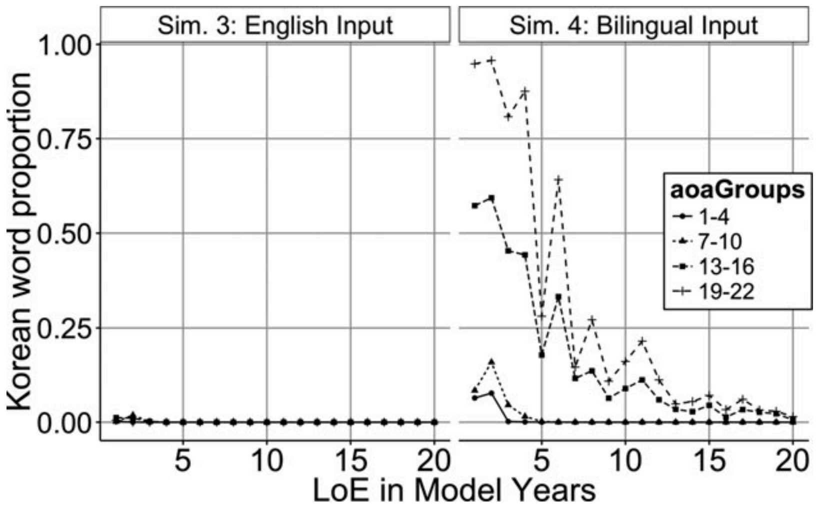


Figure 3.9. Proportion of L1 Korean words produced by English-only model and by bilingual models at different age of acquisitions (AoAs) over length of exposure (LoE).

In contrast to the marginal effect of CP in Simulation 3, the bilingual input in this simulation created a significant negative effect of CP. This means that even though the input for DET/PL was higher in the model's input, the model learned these rules less well compared to less frequent rules like 3PS/PST. We will discuss the source of these effects in the discussion. Overall, this model provided a good match to the effects of AoA, LoE, and CP seen in the Flege et al.'s reanalysis. In addition, it provided some evidence for code-switching behavior within a model of sentence production that has learned both L1 and L2.

7. General discussion

This study of L2 learning examined the interaction between AoA and input factors like LoE and CP. In support of a critical/sensitive period, our reanalysis of Flege et al.'s (1999) data found a significant effect of AoA on L2 linguistic behaviors. Some studies have argued that entrenchment with connectionist activation functions can explain sensitive period effects (A. W. Ellis & Lambon Ralph, 2000; Munakata & McClelland, 2003). Simulation 1 examined this and found that these mechanisms alone were not sufficient to explain all the features of the sensitive period in the learning of grammatical knowledge. To simulate the sensitive period effects seen in humans, we changed the model's learning rates following a stretched Z function (Granena & Long, 2013). Our claim is that this learning rate is an age-dependent learning parameter that influences L1 and L2 learning equally (some L1 phenomena can also be explained with learning rate changes, e.g., Peter, Chang, Pine, Blything, & Rowland, 2015). We can contrast this with the view that the critical period reflects specialized linguistic parameters, such as a head-direction parameter (e.g., Chomsky & Lasnik, 1993), which are set within the critical period. Instead, the use of general learning parameters here suggests that linguistic critical periods could be due to mechanisms that evolved originally for non-linguistic critical period phenomena (Knudsen, 2004; chick imprinting; Lorenz, 1937; birdsong; Marler, 1970; cochlear implants; Harrison, Gordon, & Mount, 2005).

The learning rate changes in the model may also have a role in social/motivational/input- based accounts of the sensitive period. For example, it could be the case that children receive more optimal input for language learning than adults. In order for this input to create sensitive period effects, the knowledge that is learned from early optimal input should not be overwritten by the sometimes more than 20 years of less optimal adult input. The model's stretched Z learning function is one way to ensure that early experiences due to various factors persist in spite of further learning. Thus, regardless if one believes in a purely biological account of the sensitive period, or in a social/motivational/input-based account, there needs to be an age-dependent learning mechanism that ensures that this early experience persists such that it can influence testing that takes place years later.

The main impetus for the present work was the finding that the amount of L2 input was a poor predictor of proficiency (DeKeyser, 2000; DeKeyser et al., 2010; Johnson & Newport, 1989; Lee & Schacter, 1997; McDonald, 2000). Such findings are compounded by evidence suggesting that some L2 learners are better at recognizing the grammatical use of lower frequency rules like the third person singular than higher frequency rules like determiners (Flege et al., 1999; Johnson & Newport, 1989). To explain this, we used corpus analyses to characterize the frequency of different rules (rule CP) and used this to factor out rule variation. When rule CP was added to the Flege et al.'s reanalysis, LoE went from non-significant to a significant positive effect, which suggests that the lack of LoE effects in some studies may be due to the fact that this effect was obscured by rule variation. LoE was also significant when rule was included as a factor, which demonstrates that this result does not depend on a particular approach to computing rule CPs.

We also found that late AoA learners were less sensitive to the input (LoE) than early AoA learners. Our simulation 2 showed that the stretched Z learning function was not sufficient to explain this interaction. To model this effect in simulation 3, we assigned separate learning rates to the lexical and syntactic parts of the system (Paradis, 2004; Ullman, 2001). The lexical part retained a high learning rate throughout the training, whereas the syntactic learning rate followed the stretched Z function. The early AoA models had a high syntactic learning rate, which allowed them to reconfigure their Korean syntactic representations into representations that were more appropriate for English. However, the later AoA models had a low syntactic learning rate and hence their high lexical learning rate forced them to associate English words with sequence representations that were still partially Korean. On this account, the weaker effect of LoE in late AoA learners is due to the loss of syntactic learning ability in the late learners and their greater dependence on lexical learning as a result. This account is supported by ERP studies of L2 learners' brain activity that have found that syntactic components such as the P600 differ from native learners more than lexical-semantic components such as the N400 (e.g., Hahne, 2001; Hahne & Friederici, 2001; Weber-Fox & Neville, 1996). Furthermore, recent studies have tested grammatical distinctions that yield P600 effects in native speakers and proficient L2 learners, but which yield N400 effects in some late AoA L2 learners (McLaughlin et al., 2010). Since the N400 is traditionally associated with lexical/ semantic expectations, N400 effect for a grammatical distinction supports the claim that late AoA learners may be using lexical learning to a greater degree than early AoA learners to support their syntactic processing in the L2.

Although the syntactic learning rate in the model was completely switched off at age 16, this did not fully impair the model's ability to learn syntactic regularities and to differentiate between different rules. This is because the lexical and syntactic learning rates are both being used to learn word regularities that support syntactic grammaticality judgments (e.g., DET rule depends on predicting the word the after verbs). This means that lexical and syntactic behaviors may not be transparently related to lexical and syntactic learning in human and model behavior (see the syntactic/lexical division of labor in Chang, 2002; Gordon & Dell, 2003). For example, Granena and Long (2013) argued that lexical learning ability follows a similar negative learning function as syntactic learning, but their measure of lexical learning involves multi-word collocations, which in our model would be encoded in the sequencing system and would be sensitive to the syntactic learning rate. We have shown here that lexical learning can be used to learn grammaticality constraints in a way that mimics the behavior in late L2 learners. Overall, our account predicts that under similar input conditions, early AoA learners can use their higher syntactic learning rate to learn deeper and more abstract syntactic rules than later AoA learners and support for this can be found in Hudson Kam and Newport (2005) study, which found that children were more likely than adults to regularize the artificial language that they were taught.

Although input is important for L2 learning, some L2 learners appear to perform worse with higher frequency rules like determiners than lower frequency rules like third-person singular. There was a significant negative effect of rule CP in our reanalysis of Flege et al. (1999) study and similar effects have been found in other studies (DeKeyser, 2000; Johnson, 1992; McDonald, 2000; Murakami & Alexopoulou, 2016). Since the effect is negative, it is not straightforwardly explained by input-based theories (N. C. Ellis, 2002). A likely explanation is the transfer/interference from the L1, but it is often hard to formalize the morphosyntactic similarity across languages. The fact that the model does not capture this negative relationship in Simulation 1 and 2 suggests that the separate learning rates for lexical and syntactic knowledge in Simulation 3 and 4 are important in capturing these effects. Low-frequency rules like PST and 3PS were relatively simple and the late learning models were able to correctly predict English structures with Korean syntactic representation using lexical learning to link English words with these representations. The higher frequency DET/PL/PAR rules were more complex and harder to predict from Korean representations (these rules depend more on learned syntactic knowledge). What the model highlights is an implicit assumption of transfer accounts, which is that transfer from the L1 assumes that the L2 syntax is learned slowly enough to make it preferable to link L2 words to L1 structures and this assumption is instantiated by a gradual reduction in the syntactic learning rate, whereas lexical learning rate remained high. Although we do not know the exact nature of the L1/L2 similarity that determines transfer/interference between languages, the model provides an explicit implementation of a mechanism that captures some of these transfer effects and future work should examine the nature of this mechanism and its relation to equivalent transfer effects in human studies. The models presented here are not fully realistic simulations of L2 learners. Rather, like the mixed model reanalysis, they provided a simplified representation of a complex pattern of data. It is also not the case that one simulation is the best simulation of all L2 speakers. It may be the case that early AoA learners and learners with greater LoE are more likely to be exposed to exclusively L2 input as in Simulation 3 (L2-dominant bilinguals; Flege et al., 2002), whereas late AoA learners and learners who have only a short LoE are more likely to maintain connections to their L1 as in Simulation 4 (balanced bilinguals). Furthermore, different results would arise if the same model was trained on different L1/L2 pairs (Murakami & Alexopoulou, 2016) and the present simulations do not explain variation in implicit and explicit aspects of

L2 tasks (R. Ellis, 2005; Chang et al., 2012). The main purpose of these models is to offer a starting point for developing a computational account of L2 learning.

The main innovation in the present work is the demonstration that a model of L1 language acquisition and production can explain L2 performance over various AoA, LoE, and grammatical rules. The extension to L2 learning involved minor changes in learning rates without any major architectural changes. Since the same network/mechanism is used for encoding L1 and L2 rules, the model predicts that there will be transfer between L1 and L2 structures (Foucart & Frenck-Mestre, 2012; Hartsuiker et al., 2004; Ionin & Montrul, 2010; MacWhinney, 2005; Sabourin et al., 2006) and similar brain areas/ERP signatures for L1 and L2 processing (Friederici, Steinhauer, & Pfeifer, 2002; Kotz, 2009). Learning rate variation in syntactic and lexical systems offers an account which allows the same learning mechanism and network to explain the large differences due to AoA.

Overall, this approach provides an explicit account of the complex interactions of various aspects of L1 and L2 structure learning.

Chapter 4. Rule learning in adults: The relationship between rule frequency and rule variability

1. Rationale for studies in Chapter 4

The previous chapter used a computational model to link sentence/experiment-grain learning and year-grain learning processes. It demonstrated that with minor changes in learning rate, the same computational mechanisms could model how the different numbers of years of input influence the grammatical knowledge in L2 learners. This provides support to the notion of the LAMOLL account that year-grain size learning outcomes are the result of the same learning mechanism that is responsible for experiment-grain and sentence-grain effects and that similar learning mechanisms are active in language learners throughout their lifetime.

However, an unexpected result in this study was that the performance with different grammar rules in L2 learners was negatively associated with the frequency of those rules in language input. That is the rules that occurred more frequently in corpora were learned less well by L2 learners and this is at odds with the most language learning theories, including the LAMOLL account.

It is difficult to study this effect in L2 learners because we cannot easily manipulate the frequency of particular rules that they hear over multiple years. However, given the link between the processes that take place in behavioral experiments like AGL in chapter 3 and connectionist models like the Dual Path model, it should be possible to elicit and study the negative effect of rule frequency in a non-linguistic adult AGL study. In the following chapter, we report two studies that use the AGL task developed in Chapter 2 to examine different accounts of this negative rule frequency effect to see if there is a way to explain these results within the mechanisms of the LAMOLL account.

2. Introduction

Many studies examine language acquisition by presenting adults with sentences generated from an artificial grammar (Hunt and Aslin, 2010; Saffran, 2001; 2003; Wonnacott, Newport, and Tanenhaus, 2008). It is assumed that the way that adults acquire this knowledge is similar to the way that children would acquire this knowledge from the corpus of linguistic inputs. But work in second the language (L2) learning has shown that adult language learners do not learn a language in the same way as children. A critical difference is a sensitive period, a timeframe leading up to puberty, during which language learning is the most effective (Knudsen, 2004; Lenneberg, 1967). Children seem to be more effective at acquiring language structure than adults. For example, Hudson Kam and Newport (2005) found that children were more likely than adults to find abstract rules in learning an artificial language. Thus, to better understand artificial grammar learning studies in adults, it is worthwhile to compare these results with L2 language learning studies.

Languages are learned from linguistic input, but many L2 studies have observed that length of language exposure (LoE), measured in the number of years spent in the L2 environment, does not correlate highly with linguistic knowledge of adult speakers who started learning L2 at different ages (DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Johnson & Newport, 1989; McDonald, 2000). Recently, Janciauskas & Chang (2017) showed that L2 learners' performance with different grammar rules correlates negatively with the frequency with which those rules occur in language input and this effect becomes more prominent in later L2 learners. Such findings are inconsistent with input-based language learning theories that assume that greater amount of input leads to better language-learning outcomes. It is important to understand the role of the input in L2 learning as it has implications for the understanding of artificial language learning studies. In the present study, we use the experimental paradigm developed by Janciauskas, Jessop & Chang (in preparation) to teach participants an artificial language. Critically, this artificial language has structures and rules that are similar to those in L2 language studies, allowing us to see how the properties of the input influence the learning of rules as compared to natural language learning settings.

Sensitive period effects in language are typically studied by testing groups of the second language (L2) learners who started learning the language at a different age. For example, Johnson and Newport (1989) studied Korean and Chinese speakers who migrated to the US and thus started learning English at the age ranging from 3 to 39 years. To test their knowledge of English language rules they were asked to judge the grammaticality of sentences that were either correct or violated certain grammar rule. For example, sentences like *Tom is reading a book in the bathtub* or *Tom is reading book in the bathtub* tested the knowledge of determiner rule because accepting the second sentence as grammatical meant that participants did not know that verb phrase *is reading* was supposed to be followed by the article *a* before the noun *book*. They tested the knowledge of 12 different grammar rules, examples 4 of which are presented in Table 4.1.

Table 4.1.

Examples of grammatical and ungrammatical sentences used to test grammar rule knowledge in Johnson and Newport (1989)

Rule	**Grammatical Sentence**	**Ungrammatical Sentence**
Determiner (DET)	Tom is reading a book in the bathtub	Tom is reading book in the bathtub
Plural (PL)	The farmer bought two pigs at the market.	The farmer bought two pig at the market.
Third person singular (3P)	Every Friday our neighbour washes her car	Every Friday our neighbour wash her car
Past Tense (PST)	Yesterday the hunter shot a deer	Yesterday the hunter shoots a deer

The authors found that, in addition to the overall drop in performance that is typical to sensitive period studies, participants performed better with some morphosyntactic rules than with other rules. Interestingly, those differences became more prominent as their age of L2 acquisition (AoA) increased. For example, later AoA learners showed increasingly poor performance with determiner and plural rules, while past tense and 3rd person singular rules were affected by the AoA less. Earlier studies tried to explain this rule variability in terms of intrinsic difficulty in test sentences (Bialystok & Miller, 1999), rule salience (DeKeyser, 2000), rule-based and lexically-based differentiations between test items (Flege, Yeni-Komshian, & Liu, 1999) or universal factors in learnability (Johnson & Newport, 1989). However, these explanations are more theoretical in nature and are difficult to confirm or falsify using statistical analysis or experimental techniques. Considering the number of different sensitive period theories, explaining the dynamics of rule learning at a different age in

L2 learning becomes increasingly difficult. This shows that some changes take place with age that modulate rule learning and the mechanism behind it needs to be understood.

A promising account that can explain the variability in performance seen with different rules in Johnson and Newport (1989) comes from Janciauskas and Chang (2017) study. The authors proposed that performance differences were related to the frequency with which those rules occurred in language input, which would have led L2 learners to learn some rules better than others. To provide evidence, the authors extracted frequencies (operationalized as conditional probability; CP thereafter) of different rules from English North America corpora in the CHILDES online child language database (MacWhinney, 2000) and used the extracted CPs to predict participants' performance with those rules in the Flege, Yeni-Komshian, & Liu (1999) study. The study was based on Johnson & Newport's (1989) study but had a larger sample of 240 (as opposed to 46) Korean immigrants living in the US. Surprisingly, they found that higher frequency rules were associated with worse performance. In addition, they found that rule CP interacted with AoA and length of language exposure (LoE), as measured in the number of years spent in L2 environment, showing that with increasing AoA, LoE effect on rule learning became weaker and this affected higher CP rules more than lower CP rules.

These effects are counterintuitive and somewhat controversial because they suggest that greater amount of exposure to language input was associated with worse learning outcomes. L1 theories generally assume that more frequent exposure to certain elements of language leads to better learning outcomes. For example, children learn to inflect correctly higher frequency word forms before lower frequency forms (Dąbrowska & Szczerbinski, 2006; Leonard, Caselli, & Devescovi, 2002; Räsänen, Ambridge, & Pine, 2014; Theakston, Lieven, Pine, & Rowland, 2005). Similar frequency effects have been found at various linguistic levels including phonology (Bybee, 2001, 2006; Phillips, 1984), morphology (Marchman, Wulfeck, & Weismer, 1999), and syntax (Dąbrowska & Lieven, 2005; Rowland, 2007). However, Janciauskas & Chang (2017) are not alone with their unexpected findings. For example, Gollan, Montoya, Cera, and Sandoval (2008) investigated age-related slowing effects on word use in bilingual Spanish-English speakers. They observed that with age lower-frequency words suffered less from age-related slowing. This means that with practice low-frequency words underwent a greater amount of change than higher frequency words, which is consistent with the findings in Janciauskas & Chang (2017).

Considering that language learning relies on language input, the link between L2 learners' performance and rule distribution in the input provides a way to explore this relationship using established models of language learning and processing. In their attempts to explain the negative effect of rule CP, Janciauskas & Chang (2017) modelled the result using a connectionist Dual Path model (Chang, 2002). The model had previously been used to explain a wide range of input-related behavioural data in language development (Chang, Dell, & Bock, 2006; Fitz, Chang, & Christansen, 2011; Twomey, Chang, Ambridge, 2014) in typologically-different languages (English/Japanese: Chang, 2009; German: Chang, Baumann, Pappert, & Fitz, 2014). The authors extended the model to simulate the environment where the model was trained using the Korean language for a number of years and then it was presented with English as an L2 at a different age. This simulated a typical participant in Flege et al., (1999) study.

The simulation showed that in order to explain sensitive period effects seen in adults, the model's learning rate that supported syntactic learning part of the system had to be reduced over development making late learners rely on the lexical learning part of the system to learn a language. This also flipped the direction of rule CP effect, matching the human data. The authors used this to argue that the negative rule CP effect was related to the fact that later learners could not reconfigure their Korean syntactic representations into the ones that were more appropriate for English grammar rules. However, lower frequency rules like PST and 3PS were relatively simple and late learners could predict them using representations that were still partially Korean-oriented using lexical learning to link English words with these representations. Higher frequency rules like DET/PL/PAR were more complex and harder to predict from Korean representations because they depend more on learned syntactic knowledge.

There is some debate about whether biological or some other variables that correlate with age create the sensitive period effect in L2 learning. Some theories claim that the amount of L1 and how strongly it is entrenched has an effect on the L2 learning, as opposed to purely biological age. Support for the transfer effects can be found in L2 education settings. For instance, in their review, Derakshan and Karimi (2015) found that many difficulties in L2 learning come from the similarities and differences in the structures of the two languages. If there are structural similarities, L2 learning is easier, while those learners whole L1 has little similarities of the structure face more problems learning L2. This questions the approach in Janciauskas & Chang (2017) to emphasize age-related changes in the language learning system. However, studies with feral children (Rymer, 1993), deaf children learning a sign language (Newport, 1990), or international adoptees who have no memory of L1 (Gauthier & Genesee, 2011) support the view that there is a necessary biological component in the sensitive period, supporting the complex dual nature of the sensitive period effect in the L2 learning.

While there is support for the idea of transfer effect between L1 and L2, this hypothesis would be difficult to test behaviourally because it is difficult to define what knowledge exactly is transferred from one language to another. However, there is also an alternative hypothesis that attributes the same negative frequency effects to variability in the way rules are used. For instance, Legate & Yang (2007) found that the ease or the order with which children learned verb tense morphology in English, French, and Spanish was related to the variability with which verbs were inflected in these languages.

In an unpublished version of the paper, Janciauskas & Chang (2017b) presented an additional simulation that manipulated the variability in the rule use. For example, in natural language, a higher CP rule like DET has various exceptions in its use. For instance, people have to learn when to use determiner '*a*' and when to use '*the*'. Also, determiner '*a*' has to be changed to '*an*' if the noun is singular and countable and starts with a vowel (e.g. *an uncle*). Mass nouns have optional articles (e.g. milk or the milk) and so forth. On the other hand, rules like third person singular have a much lower CP but their application is also less variable. The rule only requires that morpheme -s is added to the present tense verbs if the subject of the sentence is a singular noun. The model showed that when such variability was added to the language, it cancelled out the benefit of their high frequency in language learning, leading to a negative effect of rule CP. However, when the variability was removed, rule CP became positive, as it would be predicted by input-based learning theories. This suggests that some changes take place in older L2 learners that affects their ability to deal with rule variability as effectively as early learners do.

This finding offers an alternative hypothesis about the negative association between rule CP and linguistic behaviours to that offered in the original set of simulations. More specifically, it instantiates the hypothesis that the negative association between rule CP and linguistic behaviours arises from the higher variability associated with higher CP rule use. However, pulling different explanations apart is not possible in natural language experiments because these two processes could not be in isolation from each other.

A possible task where learning processes could be isolated and manipulated may be provided by artificial language learning studies. Janciauskas et al. (in preparation) have developed a non-linguistic artificial grammar-learning paradigm where participants could acquire English-like structures within a relatively short period of time. In this task, English sentences like “Boys like books” was expressed as letter sequence X S Q, where X stood for animate noun *boys*, Q stood for inanimate noun *books*, and S stood for *like*. Participants processed sequences in this language by selecting letters that were distributed spatially in a circle on a computer screen with a mouse. Speed and accuracy analyses showed that despite the short and non-linguistic nature of the task, participants showed many of the effects found in natural language studies, such as verb-bias, structural priming and an interaction between the two, which was predicted both by language studies and a connectionist model.

This provides a method to investigate rule learning in adult L2 learners to examine the negative effect of rule CP reported in Janciauskas & Chang (2017) in a controlled environment. The ability to test performance at different sequence positions using online measures provides the ability to detect subtle changes as language knowledge develops. Importantly, language learning can be tested in a relatively short task in the absence of social and biological changes that can accompany L2 learning. The grammar in these non-linguistic tasks allows great control over input regularities and can be modelled on real language grammar for easier comparison. We applied the method to test the hypothesis predicted by the model of Janciauskas & Chang (2017b) that rule variation could explain the negative effect of rule CP in adult L2 learners.

2.1. Morphosyntactic rule learning in an SRT task

In line with the model's prediction in Janciauskas & Chang (2017b), we hypothesized that the negative effect of rule frequency could be due to the greater variability in higher frequency rule use. We chose to investigate the learning of 4 grammar rules used in the Janciauskas and Chang (2017) study, namely determiner (DET), plural (PL), third person singular (3P) and past tense (PST). We created a language made up of letter strings that were modelled on English intransitive (e.g. the dog jumped) and transitive (e.g. the dog ate the cake) sentences. The grammatical rules were implemented as constraints on the types of sequences that were possible. For example, the plural rule in English requires inflecting plural nouns with a morpheme -s (one dog vs. two dog – s) and this was done by creating an arbitrary plural morpheme letter (e.g., G) and adding it to the noun. The sentence *two boys jump* was rendered as Q M G P, where Q is a numeral determiner *two*, M is a *boy*, G is a plural morpheme, and P is a verb *jump*. Some rules were set to occur more frequently than others and this created the differences in the rule frequency that mimicked those in Janciauskas & Chang (2017) corpus analysis. DET and PL rules had the highest conditional probability (CP), while 3P and PST rules had lower CPs.

Thus this artificial language has the basic features that are thought to create input effects across different rules in L2 language learning of English.

Since we hypothesized that rule variability was related to rule frequency, we manipulated rule variability in the language. The lower CP rules like PST and 3P were simple rules with no exceptions in their use. For example, 3PS rule always required adding third-person singular morpheme to the verb when the subject was third person singular. On the other hand, the frequent English determiner has more variability in its use. Determiners normally appeared before nouns, but they were optional with mass nouns (e.g. *I like milk*) and no determiner was required when the noun was used in generic form (e.g. *girls like boys*).

To facilitate rule learning, the study also created contexts for structural priming to occur. Structural priming is a tendency for speakers to use previously seen structures and rules (Bock, 1986). To implement this, half of the sequences that tested each of the rules were preceded by a sentence that had the same rule in that sentence position, while the other half had different rules in the adjacent sequences. While there is evidence that structural priming occurs both in first (Pickering & Ferreira, 2008; Tooley & Traxler, 2010; for reviews) and second language studies (Biria, Ameri-Golestan, & Antón-Méndez, 2010; Conroy & Antón-Méndez, 2014; Kim & McDonough, 2008; McDonough, 2006; Shin & Christianson, 2012), there is little evidence that priming effects extend to morphological rule learning. The present study will also provide an opportunity to test if repetition of morphosyntactic rules in adjacent sentences influences the processing of that rule.

Like in Janciauskas et al. (in preparation), the language was presented on a computer screen that displayed all letters of the language distributed to form a circle. Letter sequences were presented one letter at a time by highlighting the appropriate letter on the circle. Participants' task was to move the mouse from the centre of the circle to the highlighted letter as fast as possible. Each response reset the cursor back to the centre and the following letter was highlighted. Different sequences were separated by a black screen. Reaction time taken to reach the highlighted letter was recorded as the measure of their performance. Rule knowledge was tested in the sentence position where the rule was applied. For example, third person singular (3P) rule in a sequence like K H N X S (e.g. The cat like -s milk), was tested at the 4th position to see if participants expected third-person singular inflection. Based on Janciauskas and Chang (2017) results, we predicted that participants would be slower at processing higher CP rules. Based on structural priming literature, we also predicted that rules would be processed faster if the same rule was also present in the previous sentence.

3. Experiment 1

3.1. Method

3.1.1. Participants

An opportunity sample of 72 participants was recruited from the undergraduate student population at the University of Liverpool. All participants used their dominant right hand to control the computer mouse.

3.1.2. Materials

The language consisted of letter sequences that were modeled on English grammar rules. The letters belonged to one of 10 possible grammatical categories (Table 4.2) that were combined to form letter sequences that resembled English transitive (TR) or intransitive (IN) sentences (Table 4.3).

Table 4.2.

Categories and symbol grouping

Category Type	Category	Example	Symbols
Regular Noun	RNOUN	man	M, Y, H, F
Mass Noun	MASS	milk	S, D
Intransitive Verb	IVERB	jump	L, P, W
Transitive Verb	TVERB	give	N, J, T
Past tense particle	PST	-ed	Z
Plural noun marker	PL	-s	G
3rd person singular marker	TS	-s	X
Determiner	DET	a	K, R
Numeral	NUM	two	Q
Adverb	ADV	yesterday	C, B

Table 4.3.

Examples of letter sequences

Type	Category	English-equivalent example
IN	ADV NUM RNOUN PL IVERB PST	Yesterday two boy -s jump -ed
TR	DET RNOUN TVERB TS MASS	The cat like -s milk

The focus of the study was the distribution of determiner (DET), plural (PL), past tense (PST), and 3rd person singular (TS) rules. Like in the original study, rule distribution was operationalised as rule conditional probability or rule CP. DET CP was calculated by dividing the number of verbs followed by a determiner by the total number of verbs. PL CP was calculated by diving the number of nouns in plural form by the number of nouns. 3P CP was calculated by dividing the number of verbs inflected with third person singular morpheme by the total number of verbs. Finally, PST CP was calculated by dividing the number of verbs in past tense form by the number of verbs. While creating exactly the same CPs for different rules was not possible because of the smaller number of words and the limited number of constructions, the current study maintained the proportions of these distributions where DET and PL had the highest CPs and were closer to each other, while 3P and PST rules had the lowest CPs that were closer to each other (Table 4.4).

Table 4.4.

Rule CPs in Janciauskas & Chang (2017) corpus analysis, connectionist model (simulation 3) and the present study, along with the number of different instances of each rule use, as a measure of variation in experiment 1 and 2.

Rule	Corpus CPs	Modeling CPs	Present study CPs	Number of instances of different rule use	
				Exp. 1	Exp. 2
DET	0.126	0.43	0.46	5	3
PL	0.125	0.35	0.43	3	1
3P	0.05	0.14	0.29	1	1
PST	0.018	0.1	0.25	1	1

Following Janciauskas & Chang's (2017b) assumption, higher CP rules were associated with more variation in their use. Each rule is described separately below.

Determiner rule. In English language, verbs tend to be followed by nouns. Determiner rule dictated whether the noun required a determiner and what determiner was appropriate in that particular context. Depending on that, the verb was either followed by a determiner or directly by a noun. Determiners (DET or NUM) followed verbs 46% of the time. The predictability of determiner (NUM, DET) was complicated by the variability in its use. Firstly, based on whether the verb's form determiner was required either directly after a verb or after a verb and a 3P or PST marker. Also, participants had to learn how to use them in regular noun (RNOUN) or mass nouns (MASS) contexts. Table 4.4 shows possible transitions between categories. RNOUN category could be preceded by NUM or DET. However, RNOUN could also appear in generic form (e.g. 'I see birds') with no determiner. MASS nouns appeared without a determiner. However, occasionally MASS nouns were preceded by NUM (e.g. I'll have two sugars) but not by DET (e.g. I'll have a sugar) determiners.

Plural. To create the plural rule, PL category letter followed the RNOUN or MASS categories that were preceded by a numeral (NUM) that signaled plural use of noun form. The PL marker was also required when RNOUN was used in its generic form. PL marker was required in 43% of the cases.

Past Tense. To create past tense contexts, 29% of the sentences started with an adverb (ADV) denoting past time (e.g. yesterday). This required adding the past tense particle (PST) after a verb.

3rd Person Singular. The TS marker was required 25% of the times after a verb if the verb was in 'present tense' (no ADV) and the subject of the sentence (noun before the verb) was in singular form (not preceded by NUM category and was not in generic RNOUN form).

Table 4.5.

Possible transitions between categories for each rule

Rule	Transition
DET	TVERB [TS/PST] MASS
	TVERB [TS/PST] NUM MASS
	TVERB [TS/PST] NUM RNOUN
	TVERB [TS/PST] DET RNOUN
	TVERB [TS/PST] RNOUN
PL	NUM RNOUN PL
	RNOUN PL
	NUM MASS PL
TS	RNOUN TVERB TS
PST	ADV […] TVERB PST

Table 4.6 shows the first 24 sequences in the language with examples of how the four rules were instantiated. The exact sequences were generated by randomly selecting letters from appropriate categories with a bias not to choose the same letters in adjacent sequences. To test structural priming, TR structure sequences were presented in pairs, where one acted as a prime and the other was a target. Only target sequences were used to test priming. In half of the cases, the prime sequence contained the same rule that was used in the target sequence. The other half of the prime sequences had the context where the rule was omitted. This allowed testing whether rule use in the prime sequence facilitated the processing of the same rule in the target sequence. For example the sequence containing the TS rule 'DET RNOUN TVERB TS MASS' was once preceded by a prime sentence that did not require the TS rule use 'NUM RNOUN TVERB MASS' and once by a prime sentence that contained the TS rule 'DET RNOUN TVERB TS MASS'. Each prime-target pair was

separated by a filler that had an IN structure. Each 24 items made up a block with all four rules and there were 6 blocks in a study for a total of 144 sequences. These blocks ensured that each rule was tested, both with matching prime and mismatching prime, at 6 different points in development.

Table 4.6.

Example of the first 24 items in the language.

Item type	Rule Overlap	Sequence
Filler		MASS IVERB
Prime		ADV RNOUN PL TVERB PST NUM RNOUN PL
Target	Yes	ADV DET RNOUN TVERB **PST** DET RNOUN
Filler		RNOUN IVERB TS
Prime		MASS TVERB MASS
Target	No	NUM RNOUN PL TVERB NUM MASS **PL**
Filler		ADV NUM RNOUN PL IVERB PST
Prime		DET RNOUN TVERB TS DET RNOUN
Target	Yes	DET RNOUN TVERB **TS** RNOUN PL
Filler		DET RNOUN IVERB TS
Prime		MASS TVERB RNOUN PL
Target	No	NUM RNOUN PL TVERB **DET** RNOUN
Filler		DET RNOUN IVERB TS
Prime		NUM RNOUN PL TVERB DET RNOUN
Target	No	ADV DET RNOUN TVERB **PST** RNOUN PL

Filler		DET RNOUN IVERB TS
Prime		MASS TVERB RNOUN PL
Target	Yes	NUM RNOUN PL TVERB NUM MASS **PL**
Filler		ADV NUM RNOUN PL IVERB PST
Prime		RNOUN PL TVERB DET RNOUN
Target	No	DET RNOUN TVERB **TS** DET RNOUN
Filler		ADV NUM MASS PL IVERB PST
Prime		MASS TVERB DET RNOUN
Target	Yes	NUM RNOUN PL TVERB **DET** RNOUN

3.1.3. Procedure

To run the experiment we used a purpose-built Circle Task built using Processing software (version 2.2.1, www.processing.org) by Janciauskas et al. (in preparation). Standard desktop computers were used in a quiet room where up to 6 people could be tested on separate computers per session. Each participant received the instructions on a computer screen. They were informed that they would be performing a letter-matching task that they had to complete as fast and as accurately as they could. They were instructed to use a mouse with one hand while operating the keyboard whenever required with the other hand.

The experiment started with a blank screen saying ‘press SPACE to continue’. Once the key was pressed, participants saw a circle of randomly distributed letters (Fig. 4.1). After initial randomization, all participants received the same distribution of letters. At the start of the experiment, the mouse cursor was in the centre of the circles at equal distances from all letters. Letter sequences were presented one letter at a time by highlighting the appropriate letter on the circle. Participants' task was to move the mouse cursor on top of the highlighted letter as fast as possible. This reset the cursor back to the centre and the next letter in the sequence was highlighted after 200ms delay. This delay was added to provide an opportunity to anticipate the possible direction of the next letter. Letter sequences were separated by a blank screen saying 'press SPACE to continue'. Participants could either take a break if needed or continue the experiment by pressing space bar. The experiment took approximately 25min to complete.

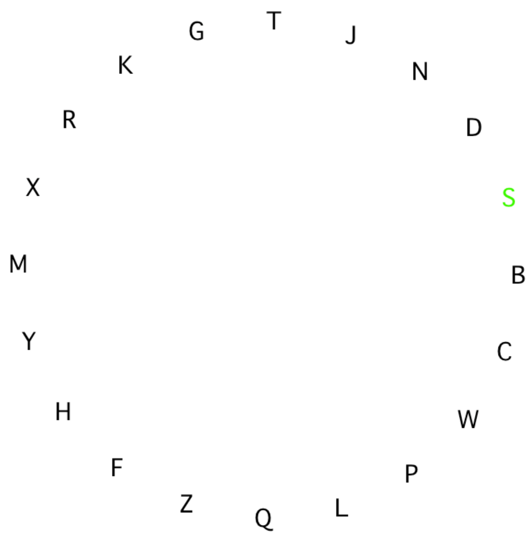


Figure 4.1. Visual display of a circle task

3.2. Results

In this and the following experiment, reaction times in milliseconds were recorded for the time taken to move the mouse cursor from the centre to the correct symbol on the circle of letters. Only correct responses were used in reaction time (RT) analyses. An error was defined as the selection of any letter that was not a target letter. The task produced a total of 55615 responses with an error rate of 12%. The data from 5 participants were excluded because of error rate higher than 50%. All rules were tested in transitive sentences after the verb. RTs were log transformed and 2 standard deviations below and above the mean were removed as outliers to normalize the data.

RTs were submitted to mixed effects linear regression analysis with test item, rule CP and prime-target match (match vs. mismatch, effects coded) as the predictor variables. Participants were included as a random effect. The model that converged included random slopes for rule CP and prime-target match (Barr, Levy, Scheepers, & Tily, 2013). P values were obtained through model comparison with likelihood-ratio tests.

The results revealed a main effect of item, showing that participants became faster overall as the experiment progressed, β= - 0.0004, *SE* = 0.0001, χ^2 = 54.3 (1), $p < .001$. This reflects general learning effects. There was a main effect of rule CP, showing that participant were slower to respond to higher CP rules, β= 0.17, *SE* = 0.04, $\chi^2(1) = 15.87$, $p < .001$. There was a significant effect of priming, where participants were faster at finding a letter when the previous sentence contained the same rule, β= - 0.01, *SE* = 0.05, χ^2 = 5.8 (1), $p = .02$. Finally, there was an interaction between item and rule CP, showing that higher CP rules improved more over the course of the study, β= -0.001, *SE* = 0.001, $\chi^2(1) = 2.87$, $p = .009$. No other main effects or interactions were observed.

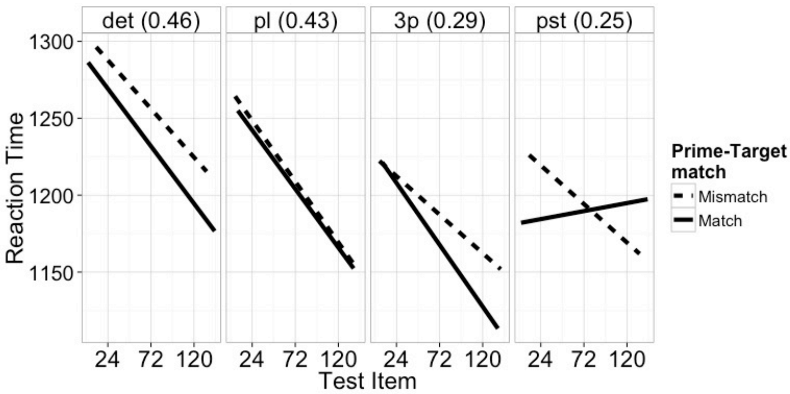


Figure 4.2. Reaction times for different rules over the course of the study when matched (solid line) and not mismatched (dashed line)

3.3. Summary of Experiment 1

The study showed that participants' speed in finding letters on the circle of letter increased significantly over the course of the study. This is consistent with the results in the original study by Janciauskas et al. (in preparation) and reflects general skill learning or practice effects. There was also an effect of rule priming in this study, where people were faster to respond when the previous sequence had the same rule in the same position. Since the prime and target sentences had different letters for the verbs and nouns, this suggests that expectations about rules were changing during the experiment in response to the distribution of sentences in the input. Thus, the frequency effects for the rules are not just due to the memorisation of previous letter sequences but are in fact due to the extraction of abstract rules. They were slower to respond to the higher frequency DET or PL rules than the lower frequency 3P or PST rules. This demonstrates that the negative correlation between rule frequency and L2 learning in English L2 acquisition can be replicated within a non-linguistic task where other factors are controlled.

4. Experiment 2

The effect of rule CP showed that participants were slower to react to the rules that had higher conditional probability. While counterintuitive, this supports the prediction of the study and is in line with the natural language results first reported in Janciauskas & Chang (2017). The direction of the effect is likely related to the greater variability associated with higher CP rule use, which was proposed in the original study. However, the second step required to test this assumption is to reduce the variation associated with higher CP rules to see if rule CP effect is reversed. In the following experiment, we have removed some exceptions associated with the higher CP rule use to makes their use less variable. At the same time, rule CPs were maintained to see how variability alone affected rule learning. Priming condition was left the same. Our prediction was that participants would perform faster with higher CP rules. Like in the previous experiment, we also expected participants to perform faster when prime and target sentences contain the same rule.

4.1. Method

4.1.1. Participants

An opportunity sample of 40 participants was recruited from the undergraduate student population at the University of Liverpool. All participants used their dominant right hand to control the computer mouse.

4.1.2. Materials

The language was created the same way as in experiment 1. The main difference was the way DET and PL rules were instantiated. Table 4.6 shows which transitions were not allowed in Experiment 2 to reduce variability in higher CP rule use. MASS noun category always occurred without a determiner. There was no variability associated with its use. Also, RNOUN category no longer occurred without a determiner and was always preceded either by DET or NUM categories. This way, all regular nouns (RNOUN) required a determiner and all MASS nouns occurred without a determiner. This reduced uncertainty associated with RNOUN use in its generic form and removed uncertainty associated with whether the determiner was required before the MASS noun category.

Since regular nouns no longer occurred in generic form and mass nouns were never preceded by NUM category, this automatically reduced the variability associated with PL rule use. PL marker was added only to RNOUN category nouns that were preceded by NUM category. The remaining two rules were identical to those in Experiment 1.

Table 4.6.

Allowed transitions between categories in both experiments

Rule	**Transition**	**Exp 1**	**Exp 2**
DET	TVERB [TS/PST] MASS	✓	✓
	TVERB [TS/PST] NUM MASS	✓	
	TVERB [TS/PST] NUM RNOUN	✓	✓
	TVERB [TS/PST] DET RNOUN	✓	✓
	TVERB [TS/PST] RNOUN	✓	
	NUM RNOUN PL	✓	✓

PL	RNOUN PL	✓	
	NUM MASS PL	✓	
TS	RNOUN TVERB TS	✓	✓
PST	ADV […] TVERB PST	✓	✓

4.1.3. Procedure

Procedure was identical to that in Experiment 1.

4.2. Results

The task produced a total of 31200 responses with the average error rate of 11.5%. Like in the previous experiment, only correct responses were used in reaction time (RT) analyses. RTs were log transformed and 2 standard deviations below and above the mean were removed as outliers to normalize the data. The data were submitted to mixed effects linear regression analysis with test item, rule CP and prime-target match (match vs. mismatch, effects coded) as the predictor variables. Participants were included as a random effect. The model that converged included random slopes for item, rule CP, prime-target match, and a two-way interaction between rule CP and prime-target match.

Consistently with the previous experiment, the results revealed that participants became faster as the experiment progressed, β= - 0.0005, *SE* = 0.0001, χ^2 = 4.84 (1), p = .03. There was a main effect of rule CP, showing that participant were slower to respond to higher CP rules, β= 0.14, *SE* = 0.05, χ^2 = 19.28 (1), $p < .001$. There was a significant effect of priming, where participants were faster at finding a letter when the previous sentence contained the same rule, β= - 0.02, *SE* = 0.007, χ^2 =

9 (1), p = .003. Finally, there was a marginal interaction between item and rule CP, showing a trend towards higher CP rules improving more over the course of the study, β= -0.001, SE = 0.0008, χ^2 = 3.44 (1), p = .06.

To compare how changes in rule variability affected the results in the second experiment, the data from both experiments were collapsed and analyzed with experiment (exp1 vs. exp2, effects coded) as an additional predictor variable. The model that converged contained random slopes for test item, rule CP, prime-target match, and experiment. In addition to the earlier seen effects of test time, rule CP and prime-target match, the results revealed that participants were faster overall in experiment 2, β= -0.05, SE = 0.02, χ^2 = 8.01 (1), p = .005. Also, the two-way interaction between test item and rule CP showed that over the course of the study participants improved more for higher CP rules, β= -0.001, SE = 0.001, χ^2 = 6.59 (1), p = .01. This effect was marginal in the previous experiment. However, while Figure 4.4 suggests small changes in DET rule and noticeable change in PL rule in the second experiment, the analysis showed no significant interactions involving rule CP between the two experiments.

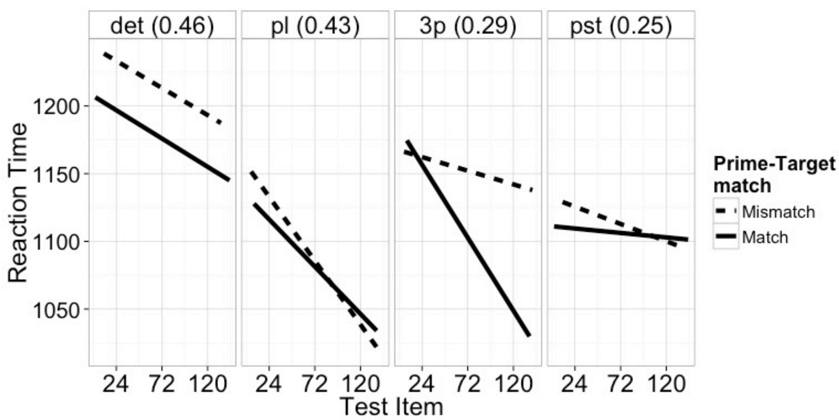


Figure 4.3. Reaction times for different rules over the course of the study when primed (solid line) and not primed (dashed line)

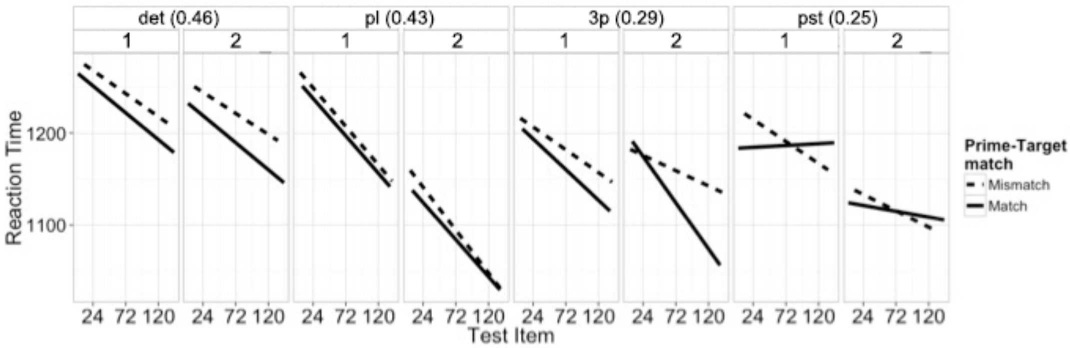


Figure 4.4. Performance with different rules in Experiment 1 and 2 when the rule was primed (solid line) or not primed (dashed line)

While there was no overall change in the effect of rule CP from reduced variability, the final analysis compared how the changes affected only DET and PL rules. The rule variable was included as a factor and was fully crossed with test item, and experiment. The model that converged included random slopes for test item, rule and experiment. The results showed that participants became faster as the experiment progressed, β= -0.0005, *SE* = 0.0001, χ^2 = 15.07 (1), $p < .001$. They processed PL rules faster overall, β= -0.06, *SE* = 0.007, χ^2 = 38.36 (1), $p < .001$. However, PL rule was also processed faster in Experiment 2 than Experiment 1, β= -0.07, *SE* = 0.014, χ^2 = 30.83 (1), $p < .001$.

The same analysis was then performed on 3P and PST rules, where no difference between the experiments was expected. The results showed that participants became faster as the experiment progressed, β= -0.001, *SE* = 0.0001, χ^2 = 19.1 (1), $p < .001$. PST rule improved less over the course of the experiment than 3P rule, β= 0.0003, *SE* = 0.0001, χ^2 = 15.2 (1), $p < .001$. Finally, PST rule was processed faster overall in Experiment 2 than Experiment 1, β= -0.04, *SE* = 0.02, χ^2 = 7.82 (1), $p = .005$.

4.3. Experiment 2 Summary

Like in the first experiment, participants became faster over the course of the experiment reflecting general learning/practice effects. The results also confirmed structural priming effects whereby repetition of the same rule in adjacent sequences improved participants' performance. The study also replicated rule CP effects seen in the first experiment and human data in Janciauskas & Chang (2017), where higher CP rules were associated with worse performance despite their higher frequency in the input. This does not support the prediction that reduced variation in higher CP rule use would change the direction of the effect. Despite the positive overall effect of rule CP, analysis of DET and PL rules, where experimental manipulation took place, showed that participants performed faster with the PL rule in experiment 2. This suggests that reduced variability in rule use affected PL rule but not DET rule. However, comparison of 3P and PST rules also showed that participants became faster using PST rule in Experiment 2, even though both rules had the same variability associated with their use in the language. Thus manipulation of DET and PL use also affected PST rule use in an unexpected way. Finally, when the data from both experiments were analyzed together, the results showed that over the course of the study participants improved more for higher CP rules. This is consistent with the general notion of input-based learning theories that higher frequency items are associated with greater learning.

5. General discussion

The goal of the study was to explain the negative rule CP effect seen in sensitive period studies of L2 learning (Janciauskas & Chang, 2017). The study explored the relationship between rule learning and rule input frequency in a non-linguistic serial reaction time task designed to study language grammar learning in experimental settings. The 25min long study demonstrated that participants became increasingly faster at finding letters on the circle as the experiment progressed. While this reflects some aspects of practice effects and learning of letter positions, performance changes were not limited to just these low-level effects. The differences in performance with individual rules showed that participants became sensitive to distributional regularities of those rules in the input and used the acquired knowledge to guide their responses. Not only were participants sensitive to rule CP but they also improved more over the course of the study processing higher CP. Such improvements are expected from the perspective of input-based language learning theories. It has been well established that children learn to inflect higher frequency word forms earlier than lower frequency words (Dąbrowska & Szczerbinski, 2006; Leonard et al., 2002; Räsänen et al., 2014; Theakston et al., 2005). Such effects have also been observed in phonology (Bybee, 2001, 2006; Phillips, 1984), morphology (Marchman et al., 1999), and syntax (Rowland, 2007) studies. However, despite better learning outcomes for higher CP rules over time, the overall performance with the same rules was worse than with lower CP rules. This is consistent with our prediction based on the study of Janciauskas & Chang (2017).

At the same time, the fact that participants improved more with higher CP rules but were overall slower when using those rules compared to lower CP rules is interesting because it suggests that both frequency and some additional process influenced participants' abilities to learn grammar rules. We proposed that greater variability associated with higher CP rule use could override frequency effects and this was supported only to some degree. When variability in rules' use was reduced for DET and PL rules in the second experiment, PL rule became easier to learn but the same was not true for the DET rule. One possible explanation is that the reduction of the variability in PL rule removed the variability from this rule altogether. This was not the case for DET rule, which had to be used in different ways in different contexts. Another possible reason is that changing variability in DET use does not affect how predictable the determiner becomes given a verb where the performance was measured. In other words, while participants can learn that DET follows verbs 46% of the time, the verb itself does not provide information about whether determiner would be required in that particular instance. Also, even if participants learned that verbs tend to be followed by the noun category letters, that knowledge alone would not provide information about whether that particular noun use requires a determiner before it unless they had information about its form before processing a verb. Some language studies have pointed out that there are differences between languages in terms of the cues that signal determiner use (Kupisch, 2007) and it may be the case that the present paradigm does not allow tapping into those aspects without some additional qualitative constraints (e.g. meaning). On the other hand, despite unchanged conditional probability for PL rule given a noun, it was easier to learn it in the second experiment because the NUM RNOUN structure was always followed by PL marker with no exceptions that were present in experiment 1. Thus variability in rule use affects at least some rule learning, which can explain some of the patterns seen in critical period studies. One unexplained observation, however, is that participants also became faster at processing the PST rule in Experiment 2, even though its use was not manipulated. The use of the rule depended on learning that sentences that started with ADV category required the use of verb with PST inflection. ADV and PST categories were always separated by two or more categories (e.g. ADV MASS TVERB PST or ADV NUM RNOUN PL TVERB PST) that involved the use of the DET rule. Simplified use of the DET rule in Experiment 2 may have made overall sequence structure learning easier, which enhanced the learning of the association between ADV and PST.

Another important finding was that rule learning/processing was influenced by structural priming. To our knowledge, this is the first study that demonstrated structural priming effects on the learning of specific morphological rules. This also extends the abilities of the paradigm to capture a wide range of phenomena that can be linked to language learning/processing. The results revealed that encountering, for example, TVERB TS transitions in the prime sentence made participants faster at finding TS category letter in the target sequence despite a total of 11 possible letters following the TVERB category and the fact that the TVERB category was made up of 3 letters itself. Since the program was biased not to repeat the same letters in adjacent sequences, the priming effect did not simply result from the repetition of identical two-letter transitions in prime and target sequences. Thus it reflected some more abstract knowledge about the categories or sentence structure and this knowledge was expressed through priming. Assuming that the present task makes use of implicit statistical learning mechanisms that natural language learning and processing rely on, it would not be surprising to find such effects in natural language learning/processing too.

In sum, the results of the present study can explain some aspects of behavioural patterns seen in sensitive period studies. Rule CP can capture participants' performance using those rules and it has been demonstrated both using corpus statistics and connectionist modelling in Janciauskas & Chang (2017) and now using behavioural measures in controlled experimental settings. It is unlikely that conditional probability itself is responsible for the observed effects because it has been established in statistical learning studies in natural language (Morgan & Newport, 1981; Saffran, Aslin, & Newport, 1996a; Thompson & Newport, 2007; Wonnacott, Newport, & Tanenhaus, 2008) and in visuo-motor learning tasks (Hunt & Aslin, 2010) that higher probability of encountering certain elements in the input is associated with better performance. It is however likely that the conditional probability reflects some other quality associated with rule use that makes the learning of those rules more difficult. One potential explanation is variability in rule use. This claim has been supported to some extent by showing that plural rule learning was made easier in Experiment 2 where its use was less variable than in Experiment 1. Similar manipulations, however, did not affect determiner rule. Thus it is likely that some other aspect that was not identified in the present study is correlated with the high frequency of rule use that makes the learning of the rule difficult. This is something for future studies to address.

5.1 Conclusion

The studies concerned with the sensitive period in L2 learning have long debated the causes that lead to the reduction in language learning abilities in older learners. Both biological and input related factors have been proposed to explain L2 learners' behaviours but there was no experimental method to tease different theories apart. In the present study, we replicated the effects associated with rule learning in sensitive period studies demonstrating that the paradigm developed by Janciauskas et al. (in preparation) offers a valuable tool to continue exploring input-related effects in greater detail. It further demonstrates that such artificial and natural language tasks make use of similar learning mechanisms, which allows testing certain language learning/processing theories in controlled experimental settings. The present study made a small first in trying to explain L2 learners' performance in sensitive period studies but more studies are needed that would look at the frequency effects in L2 learners and compare them to L1 learning.

Chapter 5: Summary of studies and general Discussion

1. Recap of the aims of the thesis

Linguistic adaptation studies show that language representations in adults change in response to the input. This process must be supported by a learning mechanism and it is important to understand how these learning processes relate to language acquisition. Language acquisition theories do not provide a mechanism that could explain how adult representations change because they often assume that language learning has a final end state and learning is no longer required once children acquire adult-like language representations.

The aim of the present thesis was to examine if language learning processes in adults and children and could be explained assuming a similar learning mechanism that is active throughout language learner's life. We refer to this mechanism as Linguistic Adaptation Mechanism of Language Learning (LAMOLL). Under this account, people make predictions about the incoming information and make small adjustments in their representations, which in turn influences how they process the incoming information. These small changes are responsible for sentence-grain linguistic adaptation effects such as structural priming, where an individual sentence changes the processing of the following sentence. However, the changes can also build up over a course of a study such that at the end of a study, the bias for one structure might be different than it was at the beginning of the study (experiment-grain changes). As more input is received, these small changes are not just responsible for adapting pre-existing structures, but can also be responsible for creating new structures. In this way, the same learning mechanism that supports linguistic adaptation can also explain year grain language learning and this was made explicit by modeling changes over L2 learning using similar mechanisms for sentence grain linguistic adaptation (structural priming, Chang, Dell, & Bock, 2006; verb bias, Twomey et al. 2014)

The present thesis reported three sets of studies that provide support for different aspects of the LAMOLL account. Firstly, the goal of these studies was to test whether the LAMOLL mechanism that is inspired by the connectionist models is psychologically real and supports learning processes in human participants. Another major aim of these studies was to provide evidence for the link between sentence-grain effects, experiment-grain and year-grain changes and show that such a mechanism was active in children and adult language learners. The rest of this chapter will summarize the findings from each study that inform different aspect of the LAMOLL account and discuss the implications and the possible directions for the future research.

2. The link between sentence-grain and experiment-gain changes

A set of 4 experiments in Chapter 2 investigated the critical assumption of the LAMOLL account that on-line processing of individual sentences is linked to sentence-grain effects like structural priming and experiment-grain effects in the form of language structure acquisition. The studies used a non-linguistic AGL task to investigate the learning of PD and DO-like structures and linguistic adaptation effects associated with them seen in natural language. Previous AGL studies used blocked designs that separated training (exposure to language) from testing (e.g. priming), which made it difficult to link these two processes. The main innovation in the present work is that the artificial language acquisition and priming were tested simultaneously during training.

The AGL studies showed that, as participants process the novel language, they slowly learned structural regularities embedded in the input and this was evident in the differences with which they processed different structures at the beginning and end of the study. These effects could be most clearly seen in terms of the faster reaction times and higher accuracy as participants processed the language over the course of each study. Structural knowledge was also indicated by the differences in participants' response to different positions of the sequences, showing that they became sensitive to contexts that predicted the possible letter transitions in each sequence position. Likewise, people showed a different response to the same category letters in different structures and the differences grew over the course of the study. This points to the growing differentiation between different sequence structures. Finally, Experiment 1 in Chapter 2 showed that people became increasingly sensitive to structural preferences of verb-like category letters, which resembled the growth of the verb bias effect over the course of the study. This further reinforces the notion that people became increasingly sensitive to the structural regularities of the language. Similar growth, however, was not obvious in the remaining experiments in Chapter 2, which suggests that methodological changes made the learning of these biases easier, in turn making it harder to detect the effect of learning. In general, these effects suggest that changes were taking place in the structural representations of the novel language over the course of the study. Such experiment-grain adaptation effects are also seen in phonology studies in natural language, where, for example, experience with certain sound patterns was shown to influence participants' speech errors over the course of the experiment, even if those patterns were not part of the language in general (Dell, Reed, Adams, & Meyer, 2000). This shows that participants readily adjust their language representations to reflect the biases in their language input.

In Experiments 3 and 4 we observed sentence-grain linguistic adaptation effects. For example, participants preferred to use the same structure that they had seen previously. Such structural priming effects grew over the course of the study, as structural representations became stronger with more language input, emphasizing the link between the sentence-grain changes and the experiment-grain changes. Additional support for the link between learning and linguistic adaptation was also provided in Chapter 4 where priming effects were evident at the level of grammar rule learning. The studies found that repetition of similar rules in adjacent sentences influenced the speed with which the rule was processed in the second sentence. Thus, we found evidence for sentence-grain effects for structure and morphological rule learning within a non-linguistic artificial grammar learning task.

This link is made stronger in Experiment 4 in Chapter 2. An important feature of the LAMOLL account, as predicted by the Dual-path model, is the idea that learning depends on prediction error, which reflects the mismatch between what the system expects and the input that it receives. The nature of such a mechanism was evident in the experiment where the mismatch between verb bias and the sentence structure of the prime sentence increased the effect of priming on the target sentence. According to the model, the mismatch in expectation to see a sequence continuation consistent with the verb's structural bias (e.g. PD structure after PD-biased verb) and the actual continuation of the sequence (e.g. DO structure after PD-biased verb) creates a larger prediction error, which in turn creates a larger change in the representations, leading to larger bias towards that structure when processing the target sequence. We found preliminary evidence for this in Experiment 4, where participants were more likely to choose prime sequence structure in their own sequence generation when the prime structure mismatched the verb's bias. These findings are consistent with a number of natural language studies that demonstrated a similar effect (Bernolet and Hartsuiker, 2010; Jaeger and Snider, 2007), but critically, these studies estimated verb-bias from linguistically-labeled corpora. Human learners do not have access to syntactically-labeled inputs and hence they need a mechanism that can acquire these regularities from the input. The Dual-path model has been shown to be able to learn verb bias without syntactically labeled inputs (Twomey et al., 2014) and hence the same prediction error-based mechanism can support both the learning of verb bias and how it is used in structural priming.

3. Year-grain changes and the link between learning in adults and children

The AGL studies provide evidence for changes that occur at short time periods between individual sequences (sentence-grain effect) and at longer time periods over the course of the experiment (experiment-grain changes). However, language learning also requires changes that persist over years. Likewise, the model's assumption that linguistic adaptation in adults reflects processes that support language acquisition hangs on a critical assumption that children and adults share a similar mechanism. While it is difficult to experimentally study language learning processes that take place over many years, it is possible to examine some of the assumptions of the model indirectly using natural experiments by looking at L2 language learning over different ages. L2 language learners receive different amounts of experience and hence it is possible to link their experience to their language knowledge.

In Chapter 3, a set of different methodologies was used to test the assumption that the same mechanism that produced short-term adaptation effects in adults was also responsible for long-term changes that resulted in structural language knowledge in children and adults. The goal of the study was to address sensitive period effects in second language learners, where older learners were shown to be less effective grammar learners than children. Specifically, the study focused on the finding that input was not correlated with older learner's L2 knowledge, as input based-learning is a critical assumption behind adaptation and acquisition. The study reanalyzed the data from Flege et al.'s (1999) study using mixed models looking at the ways that age of language acquisition interacted with people's sensitivity to the amount of language input and frequency of different grammar rules that was extracted from the language corpora. By analyzing the data from different individuals who had different amounts of experience, the mixed model created an idealized theoretical L2 learner that represented the average behaviour of the large sample of 240 L2 learners. To link this data to LAMOLL, we used a version of the Dual Path model, that was used to link structural priming to language learning in Chang, Dell, & Bock (2006), to model the behavior of this average idealized L2 learner. The same prediction-error based learning mechanism that was used to learn English structures and exhibit structural priming was used to both learn Korean as L1 and English as L2.

The Dual-Path model successfully simulated such development changes assuming the same underlying prediction-based principle of learning. The difference that affected older learners' ability to learn grammar rules was the reduced ability to rely on the syntactic learning part of the system and increased reliance on the lexical learning part of the system. This separation was instantiated in the model by lowering learning rate in the syntactic part of the system while keeping the learning rate constant in the lexical learning part of the system. The study successfully demonstrated that despite the developmental changes, the learning in children and adults could be explained assuming the same prediction based learning mechanism. In general, the fact that we were able to get a good fit to this dataset without changing the basic architecture of the model or the back-propagation learning algorithm suggests that behaviour over years in L2 learning operates on similar principles to L1 learning and linguistic adaptation.

However, it is important to clarify what we refer to as similar learning mechanisms in children and adults. The Dual-Path model successfully simulated the data, showing that learning experiences at different AoA could be explained assuming the same prediction based learning mechanism where experiences with individual sentences change language representations leading to long-term grammar knowledge. To fully model the sensitive period effects in the Flege et al. (1999) data, the L2 Dual Path model incorporated an assumption that syntactic and lexical learning language systems underwent developmental changes to a different extent. The syntactic learning part of the system became largely unavailable to late L2 learners while lexical learning remained effective in early and late learners to the same extent. In the model, this separation was instantiated as learning rate changes where learning rate was reduced significantly in the syntactic learning part of the system over the course of the development while lexical learning rate remained the same throughout the training. This separation is consistent with Ullman's (2001) theory, which separates declarative and procedural learning systems where lexical learning relies on declarative knowledge and syntactic learning relies on procedural learning. To interpret this, one could adapt the view that since the model shows that early L2 learners have full access to syntactic and lexical parts of the system and late L2 learners rely on lexical learning more than syntactic learning, the learning mechanisms are technically different and they lead to different learning outcomes. Another view is that the principles of the learning mechanism are similar across different groups and tasks. That is, regardless of the system that is engaged in learning, the learning of structure in language relies on the same principle that processing of individual sentences creates on-line changes in the language representations. In other words, the learners continuously adapt to the incoming input and we referred to this mechanism as the linguistic adaptation mechanism of language learning (LAMOLL). Both aspects have implications for drawing conclusions about the extent to which adult studies can be generalized to child language acquisition processes, a point that we will return to below.

Also, the question is to what extent the Dual Path model that simulated L1 learning and linguistic adaptation in Chang et al. (2006) and that motivated AGL studies in Chapter 2 is comparable to the L2 Dual Path model used in Chapter 3. The L1 model was used to model language acquisition and adult linguistic adaptation without additional assumptions about sensitive periods and separation of lexical and syntactic learning parts of the system. The L2 model, on the other hand specifically aimed to address the difference seen in later L2 learners in terms of their sensitivity to language exposure. However, the implementation of different learning environments was very similar in both models. Both models have a similar sequencing system, meaning, comprehension of meaning, event semantics, messages, messages are broken down in the same way. Likewise, all versions of L1 Dual Path models (e.g. Chang et al., 2006, Twomey et al., 2014) incorporate a gradual reduction in learning rate throughout the training. Essentially, these models incorporate sensitive period effects in L1 adult language simulations but these sensitive period effects are not evident because L1 users do not need to learn any new structures that could reveal the effect of the smaller learning rate in adults. The difference is that learning rate drops equally for lexical and syntactic parts of the system, as this was trivial in simulations of adult language processing. However, in theory, fixing learning rate throughout the training in L1 models would have no significant negative effect on its performance. Also, it is worth noting that in L1 model learning rate does not drop to 0, while it does in the syntactic learning part of the L2 model. However, this was done as a proof of concept rather than being a crucial aspect of the model. In theory, there is no reason why a learning rate higher than zero could not produce comparable results. Thus our claim is that in L2 models early L2 learners can be equated to child L1 learners, while older L2 learners represent adult L1 language users. The only difference is the task at hand rather than cognitive/architectural differences between them.

4. LAMOLL in the context of memory and learning

The processing-as-learning view propagated by the LAMOLL account is in line with assumptions of memory theories that see it as integrally tied to the processing systems and is both a part and a product of the ongoing processing activities of these systems (Eichenbaum & Cohen, 2004). More specifically, this link is expressed as tuning and modification of the processing networks by experience. The purpose of such adaptive processes is to maintain the efficiency of the information processing system when responding to probabilistic or deterministic regularities in the input (Bubic, von Cramon, & Schubotz, 2010). The link between linguistic adaptation and implicit learning processes is also made more explicit in Chang, Janciauskas & Fitz (2012).

In terms of the prediction-based nature of this learning mechanism, research also shows that the human brain is proactive and prediction is one of its cognitive functions (for reviews; Bar & Neta, 2008; Bar, 2009; Bubic, von Cramon, Schubotz, 2010; Gómez, Vaquero & Vazquez-Marrufo, 2004; Schubotz, 2007). In other words, the brain constantly generates predictions that anticipate the relevant future at many cognitive levels.

There is a growing body of evidence suggesting that people are making predictions in a wide variety of language-related tasks too. A range of experimental paradigms in psycholinguistic research suggests that grammatical processing is also governed by prediction-based processing systems. For example, in a speech-shadowing task, participants are required to listen to a spoken message that has to be repeated back immediately. In such tasks, the output is observed to be almost identical to input with the latencies being less than 286ms (Marslen-Wilson, 1985). However, when presented with a random word-order prose, latencies significantly increased up to 397ms. In addition, over 40% or error in shadowing the random word-order prose involved reconstruction of scrambled strings into a more standard form. Longer latencies along with the nature of errors suggest that the effect may be the result of the mismatch between the prediction and the actual structural regularities of the input.

More recent structural priming studies also demonstrated that one's attention could be biased by priming one to expect certain grammatical sequence using a visual-world eye movement paradigm (see ActaPsychologica 137 (2), 2011 issue). In a typical study (e.g. Arai, van Gompel & Scheepers, 2007; Carminati, van Gompel, Scheepers & Arai, 2008) participants hear either prepositional object dative (PO) (e.g. the pirate will send the necklace to the princess) or double object dative (DO) (e.g. the pirate will send the princess the necklace) while looking at a picture depicting all three characters. Using an eye-tracking system, they found that, when primed with the same structure sentence, on hearing the verb participants gave anticipatory looks to the recipient in the DO/DO condition and to the theme in the PO/PO condition, even though the whole structure was not yet available to guide one's gaze to the relevant object. This further supports the idea that participants were guided by the processing system's attempt to predict the structure.

Finally, evidence that grammatical learning is prediction-driven comes from event-related potentials studies. Friederici, Steinhauer & Frisch (1999) exposed participants to sentences that violated phrase structure and found an increased P600 effect reflecting the violation of anticipated structure. In contrast, Ledoux, Traxler & Swaab (2007) and Tooley, Traxler, and Swaab (2009), using ERP reported that there was a decrease in the amplitude of the P600 effect when people processed ambiguous sentences that were preceded by structurally related sentences. Taken the two studies together, the P600 effect suggests that structural violation is more costly to the system in terms of processing requirements, while the match between the system's prediction and the actual input decreases the processing cost. So there is a large body of research supporting the notion that learning in in the current studies was supported by prediction (for detailed proposal how prediction works in language see P-chain theory by Dell & Chang, 2013).

5. LAMOLL and language acquisition theories

Another interesting question is how the LAMOLL mechanism relates to the existing language acquisition theories. Child language acquisition theories offer comprehensive accounts on how children learn language structure but their proposals do not extend easily to adult learning processes. Generativist accounts assume that there is an end state for language learning once children reach adult-like levels of knowledge. Constructivist theories, on the other hand, focus on large changes that are required to learn first structures and their reliance on the mechanisms that support the learning of specific instances of structural knowledge (category learning, phrase structure, verb bias) do not allow generalizing their findings to explaining more subtle linguistic adaptation effects like abstract structural priming. The current account shifts the focus from explaining specific instances of learning (e.g. category acquisition, dative structure learning) to the characteristics of the general learning mechanism. That is, while statistical learning theories propose different mechanisms to support different levels of language learning that range from tracking transitional probabilities between sounds (Saffran et al., 1996; Onnis, Monaghan, Richmond & Chater, 2005), words (Hunt and Aslin, 2010), to transitions between categories (Saffran, 2001; 2003) and combinations of words and structures (Wonnacott et al., 2008), the current model lays emphasis on the properties of the mechanism that creates correlations between behaviours and probabilistic language regularities. Because the mechanism forms representations based on prediction, the representations naturally correlate with probabilistic regularities of the language structure. However, the model does not specifically choose what statistical regularities to collect for different levels. Rather oppositely, as the model's representations form, the model's language processing behaviour may show correlations with different probabilistic statistics of the language at different points in its development. However, in essence, the mechanism is not different to that advocated by the statistical learning theories but it lays more emphasis on the emergent nature of language representations that may differ across different populations and over different developmental stages.

In terms of its view of language, the current approach incorporates aspects of generativist and emergentist approaches. It's generativist in a sense that it has an explicit architecture that forces it to learn certain things. That is, the model will not learn regularities between words that are 5 words apart because its architecture is not designed to easily learn regularities at particular distances. Models based on simple recurrent networks have inbuilt biases that are based on the explicit assumptions about the language representations that researchers have. For example, it has a memory of a previous time step, certain hidden layer size, compression layer that forces the model to categorize things. In other words, they are built in the way that makes them best suited to learn a language in the way that we understand it. So the architecture still reflects our opinion about what language is. This way it has some similarities to some of the assumptions of generativist accounts that emphasize some language specificity.

At the same time, the model's representations are emergent in nature. In that sense, the theory is consistent with constructivist views that also assume the emergent nature of the language that treats concepts like syntactic categories as mnemonics (e.g. Ambridge, 2017). The model will form the type of representations that will help it process the language best. Because of the emergentist nature of learning, it may not be possible to clearly define the exact structure of the representations that the learner forms. However, this double-edged nature of the approach has some interesting implications for how we see the language. In this view, language is a dynamic process whose representations or expressed behaviours depend on the interaction between the learning architecture, existing representations, the learning algorithm, and the type of input. This approach is more flexible in that it sees the language process as being different in different circumstances without imposing a single categorical view of language. For instance, while early and late L2 learners engage in the same task where their behaviour is measured in using the same criteria, their language representations may be different depending on the status of lexical and syntactic systems at the time of learning, the type of representations it had at the time of learning new information (existing representations) and prediction-based learning (learning algorithm). As shown by the modelling work in Chapter 3, L2 learning behaviours in humans could be explained assuming more lexical or more abstract depending on when they started learning a second language.

6. The study of language processes using AGL

Such a dynamic nature of learning also raises a question about the extent to which behaviours that are supported by a common learning mechanism can be generalised between different studies and different populations. In other words, to what extent can adult studies inform child-learning studies and what role do AGL studies play in understanding these mechanisms? Many researchers implicitly assume that AGL studies are representative of children learning (e.g. Hunt and Aslin, 2010; Wonnacott et al., 2008) because they reflect domain-general learning. However, while the underlying learning mechanism is similar (prediction-learning or adaptation), the present L2 modeling study shows that adult learners show different learning outcomes due to their reliance on the lexical learning system. In Chapter 4 we used an AGL method to understand the negative effect of frequency on rule learning that was observed in L2 speakers in Chapter 3. The task successfully replicated the pattern of rule learning behaviour suggesting similarities in how adults extract rule regularities in artificial and natural language learning tasks. However, the model in chapter 4 also showed that the differences between different rules increase with age of language acquisition, which means that it may not be possible to use adult learning models to study language learning effects in children. This is to some extent supported by AGL studies that show that children will readily abstract rules in an artificial grammar where adults would be more reluctant to do so (Hudson Kam and Newport, 2005). The difference is also supported by Jost et al.'s (2011) visual statistical learning study that found that children showed learning-related ERP component earlier than adults, which according to the authors, suggested that they required less exposure to extract the statistical regularities from the input. At the same time, many statistical learning studies that show that results from adult and children are comparable (Saffran et al., 1999; Saffran et al., 1996b). However, the precise characteristics that make adult AGL tasks comparable to children are yet to be determined.

One of such characteristics may be the type of cognitive systems that are engaged in AGL tasks and what developmental patterns (learning rate) they follow over the course of life. As discussed above, the type of representations that learners form depend on what systems are engaged in learning, what current knowledge interacts with the learning of new information and how the two interplay with the prediction-based learning mechanism. Equating those systems to enable comparisons between different tasks and populations, however, may not be an easy task.

Just like language, AGL studies rely on many different systems, including visual that requires character recognition akin to lexical knowledge, spatial location tracking and motor system for eye movement and hand movement. How the learners represent the language may largely depend on which system is engaged in pattern finding. The four experiments in chapter 2 showed that different tasks put different demands on different systems and that has an impact on what representations people form and how their knowledge is expressed. For example, while people received comparable languages in all experiments, little structural learning was observed in experiment 1 where the language was presented in the centre of the screen. Since no spatial cues were available for the letters of the language, prediction likely involved relationships between letters rather than spatial locations. In experiment 2, 3 and 4 we observed signs of structural priming and clear verb bias effects and this improvement could be attributed to letter grouping on the circle, which could also involve more structured eye direction and hand movements. This changed how people learned representations and how behaviours were expressed where reaction times showed negative verb bias effects in Experiment 1 and positive effect in the rest of the experiments. These differences arise despite the same language and the same prediction based learning mechanism. A possible explanation of the differences may be found in Conway and Christiansen's (2009) AGL study that investigated constraints on statistical learning of spatially and temporally structured information. The authors found that participants who received information presented temporally (shapes presented one at a time in the centre of a screen) learned the language worse than those who received the sequence presented spatially (shapes presented simultaneously next to each other in the centre of the screen), while the latter was comparable to the group that received auditory information (sequences of tones of various frequencies). Thus it is important to keep in mind what systems are engaged in the task and how they compare to those in natural language tasks before we conclude how the results generalize to language learning contexts, despite similarities in learning mechanisms. The language also relies on both visual (lip reading, looking at objects) and motor (speak) systems. Likewise, sign language relies even more heavily on visual and motor system. To some extent, there are surface similarities between lexical components (perception and recognition of items) and syntactic components (motor system) in AGL and language tasks but more work is required to establish what AGL tasks provide the best match for studying language learning processes.

The AGL tasks developed in this thesis are promising in that they allow varying tasks in terms of their reliance on different cognitive systems. Language learning mechanisms have been studied for decades and many processes are difficult to study because it is not possible to isolate different components of learning in real language learning because of the rich environment in which it takes place. In some instances, only AGL could help us test the various hypothesis in isolation, like the hypothesis that variation among rules causes the negative rule frequency effect in the absence of transfer effects from different languages. Developing more precise and more controlled AGL studies could help us identify different components, how they interact, and learn the limits of a prediction-based learning mechanism. However, there is still a lot of work to be done before the precise relationship between cognitive mechanisms in these tasks is established.

7. Conclusions

Language remains immensely complex subject to study. This complexity has resulted in a split of theories in language acquisition and adult language processing theories and further specializations of different theories that focus on explaining specific phenomena of language learning or language processing. One such split is in the way language acquisition researchers view learning effects in adults as being different from learning processes in children and the way language processing theories treat the learning processes in adults without integration with language acquisition theories. The present thesis used experiments and computational modeling to link the three grain sizes. The experimental work suggested that sentence-grain linguistic adaptation changes could build up to create experiment-grain structural acquisition changes. The computational modeling work suggested that sentence/experiment grain learning mechanisms could also explain year-grain real-world language learning in L2 learners. These results are limited to a particular set of structures (dative alternation) and particular participant types (the link between sentence-grain and year-grain in L1 learners is still not clear), so more work is needed to test the generality of these results. But this work provides the first attempt to bridge across the various splits in language acquisition and adult language processing theories in order to move towards a more unified account of language.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, *23*(3), 275-290.

Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping Language Acquisition. *Cognition*, *164*, 116-143.

Abrahamse, E. L., Jiménez, L., Verwey, W. B., & Clegg, B. A. (2010). Representing serial action and perception. *Psychonomic Bulletin & Review*, *17*(5), 603–623.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Ambridge, B. (2017). Syntactic categories in child language acquisition: innate, induced or illusory? In Cohen, H. & Lefebvre, C. (Eds). Handbook of Categorization in Cognitive Science. Amsterdam: Elsevier.

Ambridge, B., Pine, J. M. & Rowland, C.F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, *123*, 260–79.

Ambridge, B., & Lieven, E. V. M. (2015). A Constructivist account of child language acquisition. In B. MacWhinney & W. O'Grady (Eds). Handbook of Language Emergence. Hoboken, NJ: Wiley Blackwell, 478-510.

Andringa, S. (2014). The use of native speaker norms in critical period hypothesis research. *Studies in Second Language Acquisition*, *36*(03), 565-596.

Arai, M., Van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*, 241–248.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*, 1235–1243.

Bar, M., & Neta, M. (2008). The proactive brain: using rudimentary information to make predictive judgments. *Journal of Consumer Behaviour, 7*(4-5), 319-330.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, *114*(3), 455–461.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language*, *11*, 557–578.

Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition, 2*(2), 127-145.

Biria, R., Ameri-Golestan, A., & Antón-Méndez, I. (2010). Syntactic Priming Effects between Modalities: A Study of Indirect Questions/Requests among Persian English Learners. *English Language Teaching, 3*(3), 111-119.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 109–127). Oxford ; New York: Oxford University Press.

Bliss, L. (1988). The development of modals. *The Journal of Applied Developmental Psychology*, *9*, 253–261.

Bloom, L. (1970). *Language development: Form and function in emerging grammars.* Cambridge, MA: MIT Press.

Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax*. The Hague: Mouton.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21*, 491-504.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, *31*, 163–186.

Bock, K., Dell, G., Chang, F., & Onishi, K. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192.

Bock, K., & Loebell, H. (1990). Framing Sentences. *Cognition*, *35*, 1–39.

Bohannon, J. N., & Marquis, A. L. (1977). Children's control of adult speech. *Child Development*, *48*, 1002–1008.

Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, *21*(5), 608–635.

Bowerman, M. (1987). Commentary: Mechanisms of language acquisition. In B. MacWhinney (Ed.), Mechanisms of language acquisition (pp. 443-466). Hillsdale, NJ: Lawrence Erlbaum.

Botvinick, M., & Plaut, D. C. (2004). Doing Without Schema Hierarchies: A Recurrent Connectionist Approach to Normal and Impaired Routine Sequential Action. *Psychological Review*, *111*(2), 395–429.

Branigan, H. P, Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition*. *157*, 250-256.

Brent, M. R. & Siskind, J. M (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, 31-44.

Brouwer, H., de Kok, D., & Fitz, H. (2013). LensOSX [Software]. Available from http://hbrouwer.github.io/lensosx/.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Bubic. A., von Cramon, Y. Y., & Schubotz. R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 22*, 1-15.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2006). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied Psycholinguistics*, *22*(02), 253–267.

Carminati, M. N., van Gompel, R. P., Scheepers, C., & Arai, M. (2008). Syntactic priming in comprehension: the role of argument order and animacy. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*(5), 1098-1110.

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, *8*(4), 322–325.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69–B77.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 821–828.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651.

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, *61*(3), 374–397.

Chang, F., Baumann, M., Pappert, S., and Fitz, H. (2015). Do lemmas speak German?: A verb position effect in German structural priming, *Cognitive Science*, 1-18.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, *90*(1), 29–49.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning: Language Adaptation. *Language and Linguistics Compass*, *6*(5), 259–278.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research, 9* (3), 198-213.

Chomsky, N. (1956). Three models for the description of language. *Information Theory, IRE Transactions on*, *2*(3), 113–124.

Chomsky, N. (1965). Aspects of the Theory of Syntax, Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1970). Remarks on Nominalization." In: R. Jacobs and P. Rosenbaum (eds.), Readings in English Transformational Grammar, Ginn, Waltham, MA, 184-221.

Chomsky, N. (1975). Reflections on Language. New York: Pantheon Books.

Chomsky, N. (1981). *Lectures on Government and Binding*. Mouton de Gruyter.

Chomsky, N. A. & Lasnik, H. (1993). The theory of principles and parameters. In J. von Stechow, A. Jacobs, W. Sternefeld & T. Vennemann (Eds.), Syntax: an international handbook of contemporary research. Berlin: De Gruyter.

Christiansen, M. H., & Chater, N. (1999a). Connectionist natural language processing: The state of the art. *Cognitive Science*, *23*(4), 417–437.

Christiansen, M. H., & Chater, N. (1999b). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157–205.

Conwell, E. & K. Demuth (2007). Early syntactic productivity: evidence from dative shift. Cognition, 103, 163-179.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In R. J. A. Sinclair & W. Levelt (Eds.), *The child's conception of language* (pp. 17-43). Berlin: Springer Verlag.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*(3), 235–253.

Clegg, B. A., DiGirolamo, G. J., and Keele, S. V. (1998). Sequence learning. *Trends in Cognitive Sciences, 2*, 275–81.

Conroy, M. A., & Antón-Méndez, I. (2014). A preposition is something you can end a sentence with: Learning English stranded prepositions through structural priming. *Second Language Research, 31*, 211-237.

Conway, C., M., & Christiansen, M., H. (2005). Statistical Learning Within and Across Modalities: Abstract versus Stimulus-Specific Representations. *In Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, 488–493.

Conway, C. M., and Christiansen, M. H. (2009). Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *Eur. J. Cogn. Psychol. 21*, 561–580.

Coyle, J. M., & Kaschak, M. P. (2008). Patterns of experience with verbs affect long-term cumulative structural priming. *Psychonomic Bulletin & Review*, *15*(5), 967–970.

Curtiss, S. (1977). Genie: a Psycholinguistic study of a modern-day `wild-child'. New York, NY: Academic Press.

Curtiss, S., Fromkin, V., Krashen, S., Rigler, D., & Rigler, M. (1974). The Linguistic Development of Genie. *Language*, *50*(3), 528–554.

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474.

Dabrowska, E., and Szczerbinski, M. (2006). Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *Journal of Child Language, 33*(03), 559-597.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(04), 499–533.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(03), 413–438.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 88-108). New York: Oxford University Press.

DeKeyser, R., & Larson-Hall, J. (2009). What does the critical period really mean. *In Handbook of Bilingualism: Psycholinguistic Approaches*, 88–108.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355–1367.

Demetras, M. (1989). *Changes in parents' conversational responses: A function of grammatical development*. Paper presented at ASHA, St. Louis, MO.

Eichenbaum, H., and Cohen, N.J. (2001). From Conditioning to Conscious Recollection: Memory Systems of the Brain. New York: Oxford University Press.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1103–1123.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(02), 143-188.

Ellis, N. C. (2013). Second language acquisition. In Trousdale, G., & Hoffmann. T. (Eds.), *Oxford Handbook of Construction Grammar* (pp. 365-378). Oxford: Oxford University Press.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227–275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*, *27*(2), 141–172.

Ellis, R. (2006). Modeling Learning Difficulty and Second Language Proficiency: The Differential Contributions of Implicit and Explicit Knowledge. *Applied Linguistics*, *27*(3), 431–463.

Ellis, R., & Wells, G. (1980). Enabling factors in adult-child discourse. *First Language*, *1*(1), 46–62.

Elman, J., L. (1990). Finding Structure in Time. *Cognitive Science*, *14*, 179–211.

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

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. Journal of Memory and Language, 158-180.

Feldman, A., & Menn, L. (2003). Up close and personal: A case study of the development of three English fillers. *Journal of Child Language*, *30*(4), 735-768.

Ferreira, V. S., & Dell, G. S. (2000). Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology*, *40*(4), 296–340.

Ferreira, F., & Clifton, C.J. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In *The 33rd annual meeting of the cognitive science society (cogsci11). Boston, MA*.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for Implicit Learning in Syntactic Comprehension. *Cognitive Science, 37*(3), 578-591.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467.

Fitz, H., Chang, F., and Christiansen, M. H. (2011). A connectionist account of the acquisition and processing of relative clauses. In E. Kidd (Eds.), *The Acquisition of Relative Clauses* (Vol 8, pp. 39–60). Amsterdam: John Benjamins.

Flege, J. E., Mackay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567–598.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104.

Fodor, Jerry A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Fodor, J., & Pylyshin, Z. (1988). Connectionism and cognitive architecture: A critical analysis. In *Connections and Symbols* (pp. 3–71). Cambridge, MA: MIT Press.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(03), 379–399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction *Psychological Review, 118*(4), 614–636.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, *99*(1), 529–534.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: a case of language acquisition beyond the "critical period." *Brain and Language*, *1*, 81–107.

Gahl, S. (2002). Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology*, *16*(12), 1173–1198.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation. *Language*, *80*(4), 748–755.

Gardell, I. (1979). *A Swedish Study of Intercountry Adoptions: A report from Allmänna Barnhuset*. Stockholm: Allmänna Barnhuset.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93.

Gauthier, K., & Genesee, F. (2011). Language Development in Internationally Adopted Children: A Special Case of Early Second Language Learning: Language Development. *Child Development*, *82*(3), 887–901.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23.

Gleason, J. B. (1980). The acquisition of social speech and politeness formulae. In H. Giles, W. P. Robinson, & S. M. P. (Eds.), *Language: Social psychological perspectives* (21-27). Oxford: UK: Pergamon.

Glüge, S., & Wendemuth, A. (2013). Solving Number Series with Simple Recurrent Networks. In *Natural and Artificial Models in Computation and Biology* (pp. 412–420). Springer.

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*(3), 1155–1164.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Gomez, C. M., Vaquero, E., and Vazquez-Marrufo, M. (2004). A neurocognitive model for short-term sensory and motor preparatory activity in humans. *Psicologica, 25*, 217–229.

Gordon, B. (1988). Preserved learning for novel information in amnesia: Evidence for multiple memory systems. *Brain and Cognition*, *7*, 257–282.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The Learnability and Acquisition of the Dative Alternation in English. *Language*, *65*(2), 203–257.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(02), 123–141.

Hall, W. S., Nagy, W. E., & Linn, R. (1984). *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale, NJ: Erlbaum.

Harris, T. & Wexler, K. (1996). The optional-infinitive stage in child English. In H. Clahsen (ed.), Generative perspectives on language acquisition, 1–42. Philadelphia: John Benjamins.

Harrison, R. V., Gordon, K. A., & Mount, R. J. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation. *Developmental Psychobiology*, *46*(3), 252–261.

Hartsuiker, R. J., & Pickering, M. J. (2008). Language integration in bilingual sentence production. *Acta Psychologica*, *128*(3), 479-489.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414.

Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, *114*(2), 151–164.

Higginson, R. P. (1985). Fixing-assimilation in language acquisition (Unpublished doctoral dissertation). Washington State University.

Hoekstra, T. & Hyams, N. (1998). Aspects of root infinitives. *Lingua*, *106*, 81–112.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482.

Huang, P. Y., Wible, D., & Ko, H. W. (2012). Frequency Effects and Transitional Probabilities in L1 and L2 Speakers' Processing of Multiword Expressions. In Stephan. T., G. & Divjak, D. (Eds.), *Frequency Effects in Language Learning and Processing* (pp. 145–175). Berlin, Germany: Mouton de Gruyter.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151–195.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31*(1), 19-31.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680.

Hunt, R. H., & Aslin, R. N. (2010). Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language*, *62*(2), 98–112.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337–374.

Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H. (2009). Dominant-language replacement: The case of international adoptees*. *Bilingualism: Language and Cognition*, *12*(02), 121–140.

Ionin, T. and Montrul, S. (2010). The role of L1-transfer in the interpretation of articles with definite plurals in L2-English. *Language Learning, 60*(4), 877-925.

Janciauskas, M. & Chang, F. (2017) Input and age-dependent variation in second language learning: A connectionist account. *Cognitive Science*, 1-36.

Janciauskas, M. & Chang, F. (2017b) Input and age-dependent variation in second language learning: A connectionist account. *Unpublished manuscript.*

Jaeger, T. F., & Snider, N. (2007). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *University of Rochester Working Papers in the Language Sciences, 3*, 26–44).

Jennings, F., Randall, B., & Tyler, L. K. (1997). Graded effects of verb subcategory preferences on parsing: Support for constraint-satisfaction models. *Language and Cognitive Processes*, *12*(4), 485–504.

Johnson, J. S. (1992). Critical Period Effects in Second Language Acquisition: The Effect of Written versus Auditory Materials on the Assessment of Grammatical Competence*. *Language Learning*, *42*(2), 217–248.

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

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946.

Jost, E., Conway, C. M., Purdy, J. D., and Hendricks, M. A. (2011). Neurophysiological correlates of visual statistical learning in adults and children. Paper Presented at the 33rd Annual meeting of the Cognitive Science Society, July 2011, Boston.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research*, *23*(6), 459–471.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic Linguistics*, 39–95.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710.

Kennison, S. M. (2009). The Use of Verb Information in Parsing: Different Statistical Analyses Lead to Contradictory Conclusions. *Journal of Psycholinguistic Research*, *38*(4), 363–378.

Kim, Y. & McDonough, K. (2008). Learners' production of passives during syntactic priming activities. *Applied Linguistics, 29* (1), 149–154.

Knudsen, E. (2004). Sensitive periods in the development of the brain and behavior. *Cognitive Neuroscience, Journal of*, *16*(8), 1412-1425.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2-3), 68–74.

Kuczaj, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16, 589–600.

Kupisch, T. (2007). Determiners in bilingual German–Italian children: What they tell us about the relation between language influence and language dominance. Bilingualism: *Language and Cognition, 10*(1), 57-78.

Lapata, M., Keller, F., & in Walde, S. S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, *30*(4), 419–435.

Lee, D., & Schacter, J. (1997). Sensitive period effects in binding theory. *Language Acquisition, 6*, 333-362.

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic Priming in Comprehension Evidence From Event-Related Potentials. *Psychological Science*, *18*(2), 135–143.

Legate, J. A. & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, *14*, 315–44.

Lenneberg, E. H. (1967). Biological foundations of language. New York, NY: John Wiley.

Leonard, L. B., Caselli, M. C, Bartolini, U., McGregor, K. K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition*, *2*(2), 151–179.

Leonard L., Caselli M. C., Devescovi A. (2002). Italian children's use of verb and noun morphology during the preschool years. *First Language, 22*, 287–304.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, R. L., Shvartsman, M., & Singh, S. (2013). The Adaptive Nature of Eye Movements in Linguistic Tasks: How Payoff and Architecture Shape Speed-Accuracy Trade-Offs. *Topics in Cognitive Science*, *5*(3), 581–610.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, *120*(11), 2546–2556.

Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In N. Ellis (Ed.), Handbook of Cognitive Linguistics and Second Language Acquisition. Lawrence Erlbaum Associates.

Lombardi, L., & Potter, M., C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*, 713–733.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*, 251-285.

Lorenz, K. (1937). The Companion in the Bird's World. The Auk, 54(1), 245- 273.

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent: Three-month-olds attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453–460.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483–506.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Third Edition. Mahwah, JN: Lawrence Erlbaum Associates.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49–67). New York: Oxford University Press.

MacWhinney, B. (2008). A unified model. In Ellis, N., & Robinson, P. (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (341-371). New York/London: Routledge.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, *5*(2), 215–234.

Marchman, V. A., Wulfeck, B., & Weismer, S. E. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech, Language, and Hearing Research*, *42*, 206-219.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*(5), 387–391.

Marler, P. (1970). A comparative approach to vocal learning: song development in White-crowned Sparrows. *Journal of Comparative and Physiological Psychology*, *71*(2p2), 1.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension.*Speech Communication, 4*, 55–73.

Mayberry, R. I. (2010). Early language acquisition and adult language ability: What sign language reveals about the critical. *The Oxford Handbook of Deaf Studies, Language, and Education*, *2*, 281-291.

Mayberry, R. I., & Eichen, E. B. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language*, *30*, 486–512.

Mayes, A. R., & Gooding, P. (1989). Enhancement of word competition priming in amnesics by cueing with previously novel associates. *Neuropsychologia*, *27*(8), 1057–1072.

Mazuka, R. (2014). The Development of Language Processing Strategies: A Cross Linguistic Study Between Japanese and English. New York: Psychology Press.

Mazurkewich, I. & L. White (1984). The acquisition of the dative alternation: unlearning overgeneralizations. *Cognition, 16*, 261-283.

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*, 395–423.

McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition, 28*, 179-207.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain Potentials Reveal Discrete Stages of L2 Grammatical Learning: Brain Potentials and L2 Grammatical Learning. *Language Learning*, *60*, 123–150.

Mermillod, M., Bonin, P., Méot, A., Ferrand, L., & Paindavoine, M. (2012). Computational Evidence That Frequency Trajectory Theory Does Not Oppose But Emerges From Age-of-Acquisition Theory. *Cognitive Science*, *36*(8), 1499–1531.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, *30*, 678 – 686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*, 91-117.

Mintz, T. H. (2006). Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), Action meets word: how children learn verbs. Oxford: Oxford University Press.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*(1), 138–153.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 213–222.

Mizumoto, T., Hayashibe, Y., Komachi, M., Nagata, M., & Matsumoto, Y. (2012). The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings. In *24th International Conference on Computational Linguistics* (p. 863-872).

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*(4), 259-305.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. Journal of Child Language, 37, 545-564.

Monner, D., Vatz, K., Morini, G., Hwang, S., & DeKeyser, R. (2013). A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition*, *16*(02), 246–265.

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8-and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447-468.

Moore, P. J. (2013). An Emergent Perspective on the Use of the First Language in the English-as-a-Foreign-Language Classroom: An Emergent Perspective on the Use of the First Language. *The Modern Language Journal*, *97*(1), 239–253.

Morgan, J., & Newport, E., L. (1981). The role of a constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 67–85.

Morisset, C. E., Barnard, K. E., & Booth, C. L. (1995). Toddlers' language development: Sex differences within social risk. *Developmental Psychology*, *31*(5), 851-865.

Mozer, M. C., & Soukup, T. (1991). Connectionist music composition based on melodic and stylistic constraints. In *Advances in Neural Information Processing Systems* (pp. 789–796).

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*(4), 413–429.

Murakami, A., & Alexopoulou, T. (2015). L1 Influence on the acquisition order of English grammatical morphemes. *Studies in Second Language Acquisition*, 1–37.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Ninio, A., Snow, C., Pan, B., & Rollins, P. (1994). Classifying communicative acts in children's interactions. *Journal of Communications Disorders*, *27*, 157-188.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*(1), B13–B23.

Onnis, L., Monaghan, P., Richmond, K., & Chater. N. (2005). Phonology impacts segmentation in speech processing. Journal of Memory and Language, 53, 225-237.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition. 126*(2), 268-284.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2007). Cumulative semantic interference as learning. *Brain and Language*, *103*(1-2), 175–176.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 786-803.

Oyama, S. (1978). The sensitive period in comprehension of speech. *NABE Journal*, *3*(1), 25-39.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Perfors, A. F., Tenenbaum, B. & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37*(3), 607-642.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*(2), 449–468.

Peter, M., Chang, F., Pine, J.M., Blything, R., Rowland, C.F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language, 81*, 1-15.

Peters, A. (1987). The role of imitation in the developing syntax of a blind child. *Text*, *7*, 289–311.

Phillips, B. S. (1984). Word Frequency and the Actuation of Sound Change. *Language*, *60*(2), 320-342.

Phillips, B. S. (2006). *Word Frequency and Lexical Diffusion*. New York: Palgrave Macmillan.

Pickering, M. J., & Branigan, H. B. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language*, *39*, 633–651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459.

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic dialogue of language. *Behavioral and Brain Sciences, 27*, 1-57.

Pine, J. M., Lieven, E. V. M., Rowland, C. F. (1998) Comparing different models of the development of the English verb category. *Linguistics 36*, 807–830.

Pinker, S. (1984). Language Learnability and Language Development. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). Learnability and Cognition: The Acquisition of Argument Structure. Cambridge, MA: The MIT Press.

Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, *20*(02), 279–303.

Potter, M., C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*, 633–654.

Potter, M., C., & Lombardi, L. (1998). Syntactic Priming in Immediate Recall of Sentences. *Journal of Memory and Language*, *38*, 265–282.

Post, K. (1994). Negative evidence. In J. Sokolov & C. Snow (Eds.), *Handbook of Research in Language Development Using CHILDES* (pp. 132–173). Hillsdale, NJ: Lawrence Erlbaum Associates.

Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). *Cognitive Development*, *11*, 19–36.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127*(2), 209–228.

Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest frequency form? *Journal of Child Language, 41*, 756-779.

Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check: E-Z Reader and event-related potentials. *Psychophysiology*, *48*(7), 993–1003.

Reitter D, Keller F, Moore JD. A computational cognitive model of syntactic priming. *Cognitive Science, 35*, 587–637.

Rescorla, L., & Reberts, J. (2002). Nominal versus verbal morpheme use in late talkers at ages 3 and 4. *Journal of Speech, Language, and Hearing Research*, *45*, 1219-l231.

Rohde, D.L.T. (1999). LENS: The light, efficient network simulator. Technical Report CMU-CS-99-164, Carnegie Mellon University, Department of Computer Science, Pittsburgh, PA.

Rollins, P. R., (2003) Caregiver contingent comments and subsequent vocabulary Comprehension. *Applied Psycholinguistics*, *24*, 221-234.

Rondal, J. A., & Anne, C. (1990). Input Evidence Regarding the Semantic Bootstrapping Hypothesis. *Journal of Child Language, 17*(3), 711-717.

Rowland, C. F. (2007). Explaining errors in children’s questions. *Cognition*, *104*(1), 106–134.

Rowland, C. F., & Noble, C. L. (2010). The Role of Syntactic Structure in Children's Sentence Comprehension: Evidence From the Dative. *Language Learning and Development*, *7*(1), 55-75.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(9), 533–536.

Rymer, R. (1993). Genie: An abused child's flight from silence. London: Michael Joseph.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Eds.), *Children's language*, Volume 4 (pp. 1-28). New York: Gardner Press.

Saffran, J. R. (2001). The Use of Predictive Dependencies in Language Learning. *Journal of Memory and Language*, *44*(4), 493–515.

Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science, 14*(4), 110-114.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996a). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., & Thiessen, E. D. (2007). Domain general learning capacities. In E. Hoff and M. Schatz (Eds.) Blackwell Handbook of Language Development, pp. 68-86. Oxford: Blackwell Publishing.

Schacter, D. L., & Graf, P. (1986). Preserved learning in amnesic patients: Perspectives from research on direct priming. *Journal of Clinical and Experimental Neuropsychology*, *8*(6), 727–743.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, *56*(2), 153-171.

Segaert, K., Menenti, L., Weber, K., & Hagoort, P. (2011). A Paradox of Syntactic Priming: Why Response Tendencies Show Priming for Passives, and Response Latencies Show Priming for Actives. *PLoS ONE*, *6*(10), 1–14.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and Phonemic Contrasts in Infants' Learning of Sound Patterns. *Language Learning and Development*, *5*(3), 191–202.

Scheepers, C., Raffray, C. and Myachykov, A. (2017) The lexical boost effect is not diagnostic of lexically-specific syntactic representations. *Journal of Memory and Language*, *95*, 102-115.

Sheng, L., Bedore, L. M., Peña, E. D., & Fiestas, C. (2013). Semantic Development in Spanish-English Bilingual Children: Effects of Age and Language Experience. *Child Development*, *84*(3), 1034–1045.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Science, 11*, 211–218.

Shin, J.-A., & Christianson, K. (2012). Structural Priming and Second Language Learning: Structural Priming and L2 Learning. *Language Learning*, *62*(3), 931–964.

Singleton, D., & Lengyel, Z. (1995). *The Age Factor in Second Language Acquisition: A Critical Look at the Critical Period Hypothesis*. Clevedon: Multilingual Matter.

Smith, M. C., & Wheeldon, L. R. (2001). Syntactic priming in spoken sentence production-an online study. *Cognition*, *78*(2), 123–164.

Snyder, W., & Stromswold, K. (1997). The structure and acquisition of English dative constructions. *Linguistic Inquiry*, *28*, 281-317.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J.L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35,* 869-902.

St Clair, M.C., Monaghan, P., & Christiansen, M.H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. Cognition, 116, 341-360.

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, *39*(3), 392–417.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103–114.

Taylor, C. F., & Houghton, G. (2005). Learning Artificial Phonotactic Constraints: Time Course, Durability, and Relationship to Natural Constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1398–1416.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics*, *16*(1), 247–277.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42.

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51–68.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209-253.

Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press.

Tooley, K. M., & Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition*, *132*(2), 101–136.

Tooley, K. M., & Traxler, M. J., & Swaab T. Y. (2009). Electrophysiological and behavioral evidence of syntactic priming in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*(1), 19-45.

Tooley, K. M., & Traxler, M. J. (2010). Syntactic Priming Effects in Comprehension: A Critical Review: Syntactic Priming Effects in Comprehension. *Language and Linguistics Compass*, *4*(10), 925–937.

Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin & Review*, *15*(1), 149–155.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *19*(3), 528.

Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, *73*, 41–71.

Twomey, K. E., Chang, F., & Ambridge, B. (2016). Lexical distributional cues, but not situational cues, are readily used to learn abstract locative verb-structure associations. *Cognition. 153*, 124-139.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model*. *Bilingualism: Language and Cognition*, *4*(1), 105–122.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten and J. Williams (Eds.), *Theories in second language acquisition: An introduction,* 2nd ed. (pp. 135-158). New York: Routledge.

Valian, V. (1986). Syntactic Categories in the Speech of Young Children. *Developmental Psychology*, 22, 562-579.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*, 21–81.

van Houten, L. (1986). *Role of maternal input in the acquisition process: The communicative strategies of adolescent and older mothers with their language learning children*. Paper presented at the Boston University Conference on Language Development, Boston.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1289–1295.

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, *112*(1), 81–96.

Warren-Leubecker, A. (1982). *Sex differences in speech to children* (Unpublished doctoral dissertation). Georgia Institute of Technology.

Weber, K., & Indefrey, P. (2009). Syntactic priming in German–English bilinguals during sentence comprehension. *NeuroImage*, *46*(4), 1164–1172.

Weber-Fox, C., & Neville, H. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Cognitive Neuroscience, Journal of*, *8*(3), 231–256.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271.

Wheeldon, L., & Smith, M. (2003). Phrase structure priming: A short-lived effect. *Language and Cognitive Processes*, *18*(4), 431–442.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua*, *106*, 23–79.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, *60*(3), 368–392.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209.

Yang, C. (2002). Knowledge and Learning in Natural Language. Oxford: Oxford University Press.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*, 1-29.