Author: Lai, Jun
Title: The learnability of center-embedded recursion : experimental studies with artificial and natural language
Issue Date: 2013-02-26

The Learnability of Center-embedded Recursion:

Experimental Studies with Artificial and Natural Language

Jun Lai

ISBN 978-90-6464-641-6

Printed by GVO drukkers & vormgevers B.V.

The research presented in this thesis was supported by a grant from the Netherlands Organization for Scientific Research (NWO)

The Learnability of Center-embedded Recursion:

Experimental Studies with Artificial and Natural Language

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus, Prof. mr. C.J.J.M. Stolker
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 26 februari 2013
klokke 11.15 uur

door

Jun Lai
Geboren te Ganzhou, China
in 1982

Promotiecommissie

Promotores:	Prof. dr. N.O. Schiller
	Prof. dr. B. Hommel
Co-promotor:	Dr. F.H. Poletiek
Overige leden	Prof. dr. P. Monaghan, Lancaster Universit
	Prof. dr. P. Zwitserlood, University of Münster
	Prof. dr. C.J. ten Cate
	Dr. W. La Heij
	Dr. Y. Chen

Contents

Chapter 1

General Introduction

From the initial stage of life, children display magnificent abilities to learn the structure of events they hear and see. From the numerous bits of information present in their environment, they are capable to learn in an efficient way, even if the information is sometimes degraded, noisy and ambiguous. For example, they quickly learn that crying is followed by mother's attention, and also that, in the phrase "doggy barks", "doggy" refers to the animal they see, and "barks" refers to what the animal does. Through experience and exposure, children can acquire even highly complex patterns and structures in various domains: e.g., concept formation (Mandler & Mcdonogh, 1993; Starkey, 1981), language comprehension and production (Schiller, 2008; Schiller & Meyer, 2003), stimulus generalization and categorization (Cohen & Strauss, 1979; Quinn, Eimas, & Rosenkrantz, 1993), action-effect learning (Eenshuistra, Weidema, & Hommel, 2004; Karbach, Kray, & Hommel, 2011), motor skill learning (Newell, 1991; Shapiro & Schmidt, 1982), and early social communication (Ayoub, Vallotton, & Mastergeorge, 2011; Helmers & Patnam, 2011).

Among all the skills of early childhood learning, one of the most prominent is natural language acquisition, and especially grammar induction (Chater & Vitanyi, 2007; Chomsky, 1957, 1965; Gold, 1967; Skinner, 1957). The question about how children perceive, comprehend and produce language in such fast paced manner has been the subject of one of the most well known debates in (psycho)linguistics since the fifties, and has intrigued researchers across various disciplines, such as psychology, linguistics, biology and philosophy (Bates, 1976; Chomsky, 1980; Christiansen & Chater, 2008; Friederici, 2004; Pinker, 1989; Tallerman et al., 2009; Tomasello, 2000). Especially, understanding the capacity to produce and understand an infinite variety of possible messages with a limited number of words and a limited set of sequential rules is still a scientific challenge.

A crucial property of language that underlies this powerful productivity is recursion. This characteristic is considered to be highly abstract and complex from a computational and cognitive point of view. It has played a major role in fundamental theoretical debates about the status of language, e.g., to distinguish humans from non-human primates, and in empirical psycholinguistic work about the learnability of complex syntax. It is against this background that the series of studies presented in this thesis have been designed. In particular, we looked at features of the linguistic input and at semantic influences that might facilitate cognitive learning and processing recursion. We assume that this learning is usage based (Christiansen & MacDonald, 2009; Tomasello, 2000) and discuss whether the learning can be explained with general learning mechanisms and working memory.

In the present introduction, the background of the thesis is sketched. First, the principle of center-embedded (CE) recursion is explained. Next, we briefly discuss animal studies on recursion learning, followed by a section with theories and experimental evidence about human learning. Here, the complexity of the principle is contrasted with pragmatic learning strategies. Then, we discuss the features of the input that might help recursion learning. Finally, we discuss the methodological issues regarding the use of artificial language to study aspects of natural language learning.

CE recursion

A recursive rule is self referential: the rule can call upon itself to form a new legal instantiation of the rule. Sentences with CE clauses in natural language are applications of linguistic recursion. For example, in the sentence "*The dog the man walks eats a bone*", the grammatical Subject-Verb-Object (SVO) construction *"the man walks the dog"* is inserted in another SVO construction "*the dog eats a bone*", making a new well formed English sentence (Fitch, 2011; Hauser, Chomsky, & Fitch, 2002). Recursion is a characteristic of almost all natural grammars (Fitch, Hauser, & Chomsky, 2005). In languages like English and Dutch, CE recursion occurs, though not frequently. Sentences with more than two levels-of-embedding (2-LoE) occur rarely in written forms of natural language, and even less in oral forms (Karlsson, 2010). Recently, however, researchers have described a language, i.e. Pirahã that has no recursive rules (Everett, 2005).

Among all varieties of recursion in language, the CE rule stands out as the focus of psycholinguistic research, because it is assumed to pose most cognitive difficulties. The reason for these difficulties is that the CE rule produces (multiple) long distance dependencies, which can not be processed in a linear way (Chomsky, 1957; Christiansen & Chater, 1999). In the English sentence "*The student that the teacher helped improved*", the sub-clause “*the teacher helped*” is inserted in the main clause “*The student improved*”. This operation results in dependencies between related components that are pushed apart from each other (e.g. “*the student*” and “*improved*”). To comprehend this sentence, the cognitive processor has to keep an initial element in memory and, further in time, relate it with its counterpart at the end of the sentence. Meanwhile, new components have to be stored in memory and bound as well. CE recursion requires a high level of mental processing; both in terms of memory and computation (Gibson, 1998).

The learnability of recursion has not only evoked an intensive theoretical debate on the evolution and the status of language, but has also spurred behavioral studies with human and non-human species (Gentner, Fenn, Margoliash, & Nusbaum, 2006; Hauser et al., 2002; Lai & Poletiek, 2011; Rey, Perruchet, & Fagot, 2012). Here, the main aspects of this debate and related data are summarized. A major question in the debate about recursion in language is whether it explains the borderline between human and non-human communication systems, and how it has emerged in the evolution of human language. Regarding the evolution of recursion, there are two views: the “saltationist” and the “gradualist” view (Coolidge, Overmann, & Wynn, 2011). The saltationists regard the emergence of recursion as a “genetic change”, which is adaptive to non-language related functions (Reuland, 2010). In a seminal paper, Hauser et al. (2002) proposed a “recursion-only” framework (Pinker & Jackendoff, 2005), in which they define recursion as a unique attribute of language, which could distinguish the faculty of language in the broad sense (FLB) from the faculty of language in the narrow sense (FLN). The key difference between FLB and FLN is proposed to be biologically-based in the sense that FLB is common to both human and non-human primates, while FLN is available uniquely to human beings (Corballis, 2007; Friederici, 2004). Hence, the saltationists regard the emergence of recursion to be all of a sudden and they propose that FLN, which includes recursion as the crucial distinctive component, may have emerged for purposes other than communication, such as navigation, social interaction, etc.

On the contrary, the gradualists indicate that recursion emerged gradually and that the evolutionary purpose of language actually is aimed for communication (Coolidge et al., 2011). For instance, Pinker and Jackendoff (2005) posed a strong opposition to the “recursion-only claim” by stating that the saltationists overweighed the recursive component of human language, overlooking other non-recursive aspects, such as phonology and morphology, which are also unique to human language. Gradualists dispute the theory that recursion-only underlies the distinction between human and animal communication systems, pointing at various other non-syntactical characteristics of human language that have changed gradually along with the evolution of the human species.

Can birds and monkeys learn CE recursion?

The debate on the origin of human language was boosted by findings from studies with non-human species. Animal studies on recursion have investigated two main questions. First, does the ability to process the specific CE structures belong uniquely to human beings or not?; Second, *if* animals show the ability to process CE, does the performance reflect true detection of CE structures, or does it merely reflect the application of simple substitute strategies? The findings are far from conclusive (Beckers, Bolhuis, Okanoya, & Berwick, 2012). For instance, Fitch and Hauser (2004) showed that cotton-top tamarins were only able to learn an artificial finite state (linear) grammar, but not a recursive phrase structure grammar, while human beings could learn both grammars. Fitch and Hauser therefore proposed that this result indicates that the ability of processing CE recursion distinguishes humans from nonhumans.

In a recent experiment, however, Rey, Perruchet and Fagot (2012) showed that after having been trained on a basic structure of two elements, baboons preferred new sequences with two combined basic structures in one sequence, which were ordered according to a CE structure, over sequences following any other structure. The authors conclude that CE structures may have evolved under the influence of very low level mechanisms, shared by humans and baboons. The conclusion that the baboons’ responses are related to evolutionary pressure favoring CE constructions in human languages has been doubted, however (Poletiek & Fitz, submitted). Thus, though it is unclear to what extent non-human primates can “parse” long distance dependencies, in some studies, their behavior superficially correlates with knowledge about distant elements depending on each other.

Findings from bird studies also challenge the uniqueness of recursion to humans. For instance, Abe and Watanable (2011) first detected that Bengalese finches show a robust sensitivity to complex syntactic structure with non-adjacent dependencies that were generated by an artificial grammar. Successively, research of Bloomfield, Gentner and Margoliash (2011) suggested that songbirds may skillfully use statistical information in their environment to help themselves in learning long-distance matches. Analogously, European starlings were found to show recognition and discrimination between linear and embedded structures (Gentner et al., 2006). However, as in studies with primates, there is no consensus over the exact "knowledge" that songbirds use when processing center-embeddings (Berwick, Beckers, Okanoya, & Bolhuis, 2012; Coolidge et al., 2011; Corballis, 2007; Friederici, 2012; Rey et al., 2012). For instance, van Heijningen, de Visser, Zuidema, and ten Cate (2009) showed that zebra finches (seven out of eight) were able to distinguish 1-LoE CE structure. However, the finches failed to generalize this recursive rule to new items with the same structure (e.g. AABB) that came from another domain of elements (e.g. CCDD). The only bird, which successfully transferred the distinction across item categories, was later shown to be using other simple heuristics than the hierarchical structure. Generally speaking, songbirds may apply cognitively simple strategies in matching acoustic similarities that apparently coincide with the recursive rule to perform the experimental task (Beckers et al., 2012). It might not be the actual abstract hierarchical recursive principle that was learned, but the mere regularities that looked like or could be described computationally as recursive CE.

Summing up, animal studies on recursive learning suggest that some non-human beings might have the capability to learn a CE pattern. However, this capacity is limited to 1-LoE and vocabulary learning is limited as well. Moreover, the actual observed performance by animals in these studies could mostly be attributed to superficial mechanisms instead of actual knowledge of the hierarchical positional pattern of recursive CE. These limits make it problematic to interpret animal performance in terms of "learning recursion". The ambiguous findings about the learnability of recursion by animals, now, raise the question how humans actually process CE recursion. Do humans learn more and process more deeply CE structures in the context of language learning and language use than animals? In other words, do they reach the essentially higher stage of knowledge that was referred to by Hauser et al. (2002) as FLN? Or are the learning processes and the usage of these types of hierarchical structures limited in the same way as animal learning seems to be (Perruchet & Rey, 2005)? After all, these structures are, also for humans, quite hard to process (Abney & Johnson, 1991; Anderson, 1976; Baum, 1993; Christiansen & MacDonald, 2009; de Vries, Petersson, Geukes, Zwitserlood, & Christiansen, 2012; Lai & Poletiek, 2010; Schlesinger, 1975; Weckerly & Elman, 1992). What explains these difficulties and how do language users overcome them?

Human processing of CE recursion

There are various theories accounting for the parsing difficulty caused by complex CE recursive structures: for instance, *the processing overload theory* (Gibson & Thomas, 1996; Kimball, 1973; Lewis, 1996) points at the limited cognitive abilities such as working memory capacity. Long-distance dependencies consume more resources when associating corresponding elements, than linear right-branching (RB) recursion. Gibson (1998) pointed at two kinds of costs in processing CE recursion: first, integration costs, which are enhanced along with the increase of distance and number of related elements; second, memory costs, which are used for storing all information until the whole structure is terminated.

The *structural configuration theory* (Chomsky, 1965; Johnson, 1998; Miller & Isard, 1964) explains processing difficulties by how CE structures are constructed. To process CE recursion, human parsers solve a complex puzzle: they need to relate elements, which “are bound from the outside in” (Corballis, 2007) and discover where the new embedding starts. Finally, some researchers have explained the difficulties from a purely logical point of view. The *incomplete dependency account* (Johnson, 1998) perceives the difficulty as “geometric constraints” of a proof net. The breakdown of processing occurs when there are too many unsatisfied relations (too many A’s in memory waiting in vein for a B to be paired with, to clarify the semantic content of the sentence) (Morrill, 2000). Studies from the field of discourse analysis refer to this problem as “unfinished thematic dependencies” (Hakuta, 1981; MacWhinney, 1987; Pickering & Barry, 1991).

Since CE recursion is so difficult to process by humans, while even animals seem able to recognize aspects of the CE structure superficially, what exactly do humans know about these structures when they use or "parse" them? What knowledge is recruited to solve the CE puzzle? Research on CE recursion learning with the artificial grammar learning paradigm (AGL) shows that several degrees of "abstractness" of knowledge about CE can be distinguished (de Vries, Monaghan, Knecht, & Zwitserlood, 2008). Before presenting the results of this research, we first describe the AGL procedure, and the experimental grammar stimuli used in this paradigm to test CE structures.

In AGL, a participant is first exposed to exemplars of the grammar without any explanation about the rules underlying them. This grammar learning by mere exposure simulates the situation in which a child is exposed to linguistic utterances. In the subsequent test phase, participants would be tested with new sequences, half of which are grammatical and half ungrammatical. Participants give grammaticality judgments for the test items, judging whether they are governed by the same rules as the ones underlying the training items. To analyze the knowledge involved in CE processing in a lab context, typically, a reduced version of a CE grammar is used, called A^nB^n structures (Fitch & Hauser, 2004). This grammar has two word categories (A-words and B-words, for example, referring to nouns and verbs respectively in natural language). The basic structure of the grammar is a string AiBi, in which a particular A-word can be legally associated with a particular B-word according to the basic rules of the grammar. The recursive CE operation involves insertion of a grammatical AjBj string within an AiBi string, resulting in a grammatical string AiAjBjBi. This insertion operation can be applied an infinite number of times, resulting in an infinite output set of grammatical sentences.

First, one of the most superficial characteristics of an A^nB^n grammar is that a grammatical sequence should have an equal number of A's and B's. If this rule is learned only, distinguishing grammatical from ungrammatical sequences would boil down to counting A's and B's. De Vries et al. (2008) found that participants in an artificial grammar learning task could easily learn this feature of a CE rule. Another superficial characteristic of CE can be induced from exemplars with repeated words. For example, repeated A-word in the beginning of an $A_1A_2B_2B_1$ structure (A_1A_2 being the same word) and repeated B-words provide a strong cue that A-words are different from B-words, and that the equality of the A-words might be related to the equality of the B-words. Learners focusing on this feature might judge the grammaticality of a new sentence, by checking whether the B-words are grammatically related to the A-words, without any consideration of the sequential order of the B's. The use of these superficial characteristics has been found in several studies on CE processing (see e.g. Rohrmeier, Fu, & Dienes, 2012, for a review). Indeed, previous studies, which suggest that participants could recognize the A^nB^n type of sequences (Bahlmann, Gunter, & Friederici, 2006; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006), used test items such as AAAB (Bahlmann et al., 2006) and AABA (Friederici et al., 2006) that could easily be detected as ungrammatical without any knowledge of the CE rule, merely by counting and checking the numbers of A's and B's and the transitions from A to B words (that was only permitted in the middle of a grammatical sequence). Hence, the knowledge acquired and used to process CE sentences might correlate with, but not cover the full complexity of the CE structure. Overt behavior in a particular experimental task may look like it is reflecting abstract CE recursive knowledge, but in fact may be based on superficial aspects of it. A substantial part of the observations on the learnability of CE hierarchical structures, with both human and non human species, might be the visible result of superficial task dependent strategies, not hierarchical processing per se (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Corballis, 2007; de Vries et al., 2008).

In response to this problem, some experimental work focusing on hierarchical processing has been conducted, attempting to exclude as much as possible superficial strategies. For instance, in an fMRI study, Bahlmann, Schubotz and Friederici (2008) used two types of artificial grammars, i.e. A^nB^n and $(AB)^n$, and assigned a CE versus a RB mapping between A- and B- categories (e.g. $A_1A_2B_2B_1$, or $A_1B_1A_2B_2$). They found higher brain activities in Broca's area when participants processed A^nB^n rather than $(AB)^n$. In another study, de Vries et al. (2008) used the same training materials in a behavioral study and manipulated the type of violations in the test items. They introduced scrambled ungrammatical items (e.g. $A_1A_2A_3B_1B_3B_2$), which they considered to be the most difficult violation to detect, and only detectable with full knowledge about all aspects of the CE structure. Their participants failed to distinguish the ungrammatical items. However, when the scrambled ungrammatical items contained an easy feature to detect as well, like syllable repetitions, participants showed above chance performance. Therefore, de Vries et al. concluded that there was no evidence supporting real learning of CE recursion in AGL.

Two additional possible experimental procedures have been used to test “true recursion” in AGL. First, test whether participants can generalize CE rules to a higher level than they have been exposed to during training (Poletiek, 2002). As Poletiek (2002) notices, however, adding one LoE in test items possibly increases memory load. When participants fail to parse these longer items correctly, which they did in Poletiek’s study, this may be due to memory limitations rather than to the actual incapability to generalize the recursive operation to higher levels of complexity. Second, deep processing of CE might be investigated by testing transfer of knowledge: can participants transfer their knowledge of the CE rules to novel items which are containing the same structures, but contain elements from another domain (Kinder, Shanks, Cock, & Tunney, 2003; van Heijningen et al., 2009)? Studies thus far provide mixed evidence for this capability. Indeed, there is no unambiguous evidence that participants use the actual CE rules when transferring their knowledge from one domain to another. For example, detecting repetitions has been shown to be a heuristic in transfer tasks (Redington & Chater, 1996). Interestingly, this repetition monitoring is exactly what van Heijningen et al. (2009) found zebra finches did in a transfer task.

If learning, processing and producing hierarchical CE structures is hard, occasionally even so hard that language users may have recourse to pragmatic solutions like heuristics to learn and parse them, are there maybe conditions independent of the language stimuli themselves present in the learning environment, which might help this learning?

Factors in the language environment facilitating CE processing

Processing CE recursive structures has been shown to improve under various circumstances. For instance, *the starting small approach* was initiated by Elman (1991, 1993), who observed that a simple recurrent network (SRN) showed better learning when trained piece by piece with the input, instead of being trained with the whole input at once.

A number of studies verified the facilitation effect of staged input (Cochran, McDonald, & Parault, 1999; Conway, Ellefson, & Christiansen, 2003; Kareev, Lieberman, & Lev, 1997; Kersten & Earles, 2001; Newport, 1990; Plunkett & Marchman, 1993). Particularly, Lai and Poletiek (2011) (Chapter 2 and 3 of the present thesis) found a facilitation effect of starting-small in an AGL study. In the same study, another strong positive effect on learning recursive structures was found: Extensive and early exposure to simple adjacent AB pairs *without* any embedding made detection of the CE structure much easier.

A third helpful condition for detecting CE structure is the frequency distribution of the input items, per level of complexity (see Poletiek & Chater, 2006, for a study with a non-recursive grammar). In a mathematical analysis (Poletiek & Lai, 2012) we argued that learning is helped with unequal frequencies, i.e. skewed learning distributions of items favoring high frequencies for short and simple structures. Poletiek and Lai (2012) argue that this statistical effect reflects a semantic bias effect in natural language. Indeed, the gist of the frequency effect is that some AB pairs are more frequent than other ones in the linguistic input. For example, "*dog barks*" will be encountered more frequently than "*girl barks*", and this difference in occurrence might serve as a cue for relating A's to B's: in the sentence *the dog the girl walks barks*, the difference in frequencies between *dog barks* and *girl barks* is a cue for associating *dog* to *bark* rather than *girl* to *bark*. A number of studies with natural language (Blauberg & Braine, 1974; Fedor, Varga, & Szathmary, 2012; Rohde & Plaut, 1999; Stolz, 1967; Weckerly & Elman, 1992) show that when CE sentences contain semantic biased components, the performance of sentence parsing is significantly improved compared to the situation with semantically neutral, unbiased equally frequent word pairs (Powell & Peters, 1973).

Finally, various other types of statistical information in the input seem to help exploring sequential structures (Gennari & MacDonald, 2008; MacDonald, Pearlmutter, & Seidenberg, 1994; Reali & Christiansen, 2007). For example, variations in variability of words in adjacent positions have been shown to be informative (as in "*he is working*", "*is*" and "*–ing*" are constant whilst the middle morpheme highly varies) (Gomez, 2002). Mintz (2003) has proposed a similar statistical effect in the "frequent frames" model. This distributional model could successfully predict the categorization of a target word *x* in the structure of *A_x_B*, in which *A* and *B* co-occur frequently. Finally, enhancement of intelligibility of CE recursion has been shown to be affected by other cues, such as the nouns' animacy cues (Mak, Vonk, & Schriefers, 2002, 2006), and prosodic cues (Mueller, Bahlmann, & Friederici, 2010). Hence, a number of extra linguistic aspects of the sample of stimuli that a learner is presented with seem to facilitate substantially learning complex structure. These factors together with general learning mechanisms might interact to eventually obtain knowledge of CE structures. This possibility is the focus of the present work.

Artificial or Natural Language Experiments?

In the research reported in the present thesis, mostly artificial materials have been used in laboratory experiments, with one exception (Chapter 4) using natural language sentences. Typically, in AGL research, the experimental procedure is considered to simulate the situation of a child learning natural language, reducing the natural learning period to the duration of one experimental session, and adapting the system to be learned from a full human language to an extremely simplified grammar made up of only a few non-words and only those rules that are the focus of the experimental test. Here, the type of rule that we focus on is the CE A^nB^n grammar.

Using the AGL paradigm (Reber, 1967, 1989), experimenters can manipulate the stimulus set and the features of the learning situation to study specific influences on the learning process in isolation. For example, besides rule structure, the effect of small versus large learning sets, feedback during learning and noisy versus fully correct learning input can be manipulated (Gomez & Gerken, 2000). Since a few decades, the AGL paradigm has indeed been widely used to study language acquisition and grammar induction processes (Johnstone & Shanks, 2001; Knowlton & Squire, 1994; Lobina, 2011; Marcus, Vijayan, Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996). Though, at first sight, the absence of semantics seems a drawback of AGL for generalizing results to natural grammar acquisition, this may also be seen as its strength. The semantic richness of natural language makes it hard to separate semantic and syntactic effects on language learning. Also, disregarding semantic influences makes results of AGL research comparable with machine learning performance (e.g. the SRN), which is necessarily tested on restricted and meaningless input samples (Christiansen & Chater, 1999; Elman, 1991, 1993; Misyak, Christiansen, & Tomblin, 2009; Rohde & Plaut, 1999). Besides behavioral data, AGL is also used in collecting neuroanatomical data from fMRI experiments focusing on brain activity related to syntactic processing only (Bahlmann et al., 2008; Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Friederici, Bahlmann, Friedrich, & Makuuchi, 2011; Makuuchi, Bahlmann, Anwander, & Friederici, 2009). Finally, for testing the particular status of CE syntax in the human language faculty, the advantage of AGL as a pure test of syntactic processing is particularly suitable (Poletiek, 2002; Udden et al., 2009), because the focus of the arguments is on the complexity of the grammar.

Nonetheless, the artificial nature of the AGL paradigm poses limitations to its use as well. For example, the highly positive effect on learning CE recursion of early intensive training with simple sentences *without* recursion found in our AGL study (Chapter 2) may be argued to generalize to the natural situation, where child directed speech input is also made of simple basic sentences. We do not know, however, how the semantic content of this early input interacts with early simple-structure learning. Hence, AGL is obviously limited in the sense that the full richness of the environment is not reflected. The question to what extent this limits the representativity of the results for learning outside the lab will depend on the particular goal of a study. For each experimental result, the ecological validity of the paradigm needs to be accounted for (Arciuli & Torkildsen, 2012). To investigate semantic influences on learning CE, studies with natural language materials are needed. Therefore, in Chapter 4, we present an experiment with natural language materials, in which the semantic congruency between syntactic and semantic features of CE sentences is manipulated.

Outline of the dissertation

The dissertation consists of the present introduction to the topic, four chapters reporting empirical studies, and a summary chapter. The chapters are based on manuscripts that are currently published (Chapter 2), in press (Chapter 3), under revision (Chapter 5), or submitted (Chapter 4).

Chapter 2 reports an artificial grammar learning study investigating whether the acquisition of hierarchical CE structures could be enhanced if the ordering of the learning input is staged. Participants were exposed to 144 non-sense Consonant-Vowel-syllable strings, generated by a phrase structure grammar, in an AGL task. They delivered grammaticality judgments over 144 novel strings, which were either in accordance with the same underlying rule, or were ungrammatical, i.e. violations of the rule. Results of the two experiments suggest that participants could only perform significantly above chance level performance, under two conditions: First, the input should be presented in a *starting small* fashion; and second, early learning of the basic structure of the grammar, the adjacent-dependencies is needed before the embedding structure is presented. Besides replicating the classic starting small effect (Elman, 1991, 1993), our study uncovers, for the first time, that early acquired robust knowledge of the *basic structure* of a hierarchical CE grammar is a prerequisite for subsequent acquisition of the full complex hierarchical embedding pattern later on.

Chapter 3 further explores the starting small effect in processing recursive CE structures. Specifically, this study focuses on two variants of the starting small organization of the input: on the one hand, the discretely growing input as implemented in Lai and Poletiek (2011), in which the sentences are clustered according to the number of LoE they have (first 0-LoE sentences only, next 1-LoE items only, and finally 2-LoE items only), and on the other hand, a gradually growing input (with more complex sentences being added to the stimulus sample presented over time). A second manipulation was the frequency distribution of the input sentences. We compared equal frequencies for all LoE items, with a skewed distribution in which more stimulus items of the lower LoE were presented. The results of the two experiments showed that the gradual starting small ordering was helpful only if accompanied by a skewed frequency distribution. In other words, gradually inserting more complex sentences only helps if there are much more simple basic sentences than embedded sentences in the training input. This combined effect of gradual starting small and skewed frequencies reflect the properties of the natural language input, as we argue. That input in natural language is skewed in the same way as in our AGL study, though to a more extreme extent (Kurumada, Meylan, & Frank, 2011). Moreover, complex constructions with relative clauses typically are absent in child directed speech before the age of 5 (Kidd, Brandt, Lieven, & Tomasello, 2007).

Chapter 4 uses natural language materials, and aims to make a connection between AGL studies and natural acquisition of complex recursive structures. The current study compares processing Dutch RB embedded sentences ($(AB)^n$) with CE sentences (A^nB^n). We tested the influence of the congruency between the semantic pattern of relations and the syntactic pattern of relations between the nouns (reflecting A-words) and the verbs (B-words) in a sentence. The semantic pattern could either match or mismatch the syntactic pattern, as in the sentences *The girl the dog bites cries*, and *The dog the girl bites cries*, respectively. The results showed a facilitative effect of semantic-syntactic congruency and we proposed a semantic-memory model for processing recursive (SMR) structures to account for this effect.

Chapter 5 further tested the starting small effect with different types of recursive structures and different types of staged input. In Experiment 1 and 2, we observed a facilitation effect of starting small in parsing two types of recursive grammars: RB and CE. However, sentence complexity (i.e. LoE) and sentence length were confounded in the input. Indeed, thus far, the starting small learning condition in experimental research features an ordering of sentences along two perfectly correlated dimensions: the (increasing) number of LoE and sentence length. For example, the grammar used in the study in Chapter 2, produces sentences with 0-LoE having all two syllables, 1-LoE items having four syllables, and 2-LoE items with six syllables. In Experiment 3 we disentangled these two factors, and found that participants showed learning only when the input was arranged according to complexity (LoE), and not when it was organized according to sentence length. The results suggest that the starting small input is effective because it helps learners to detect structure, not because it reduces memory load in the earliest stage of learning.

Chapter 2

The Impact of Adjacent-Dependencies and Staged-Input on the Learnability of Center-Embedded Hierarchical Structures

---

This chapter is based on: Lai, J. & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118(2),* 265-273.

Abstract

A theoretical debate in artificial grammar learning (AGL) regards the learnability of hierarchical structures. Recent studies using an A^nB^n grammar draw conflicting conclusions (Bahlmann & Friederici, 2006; de Vries, Monaghan, Knecht, & Zwitserlood, 2008). We argue that 2 conditions crucially affect learning A^nB^n structures: sufficient exposure to zero-level-of-embedding (0-LoE) exemplars and a staged input. In 2 AGL experiments, learning was observed only when the training set was staged and contained 0-LoE exemplars. Our results might help understanding how natural complex structures are learned from exemplars.

Recursion, as in sentences with hierarchically built up center-embeddings, is regarded as a crucial property of human language (Hauser, Chomsky, & Fitch, 2002). However, sentences with several levels of embedding (LoE) are difficult to process, even for native speakers (Bach, Brown, & Marslen-Wilson, 1986; Hudson, 1996; Newmeyer, 1988; Vasishth, 2001). *The rat the cat the dog chased killed ate the malt* (Chomsky & Miller, 1963, 286-287) is a typical center-embedded sentence incorporating two sub-clauses. The dependencies between related constituents become harder to associate as more clauses are inserted, not least since the counterparts get further away from each other.

Recursion refers to structures that are self-referential, and infinitely productive. In center-embedded structures, inserting a grammatical sentence within another generates a new grammatical sentence. This operation can be applied infinitely, generating numerous output sentences. Since Hauser, Chomsky and Fitch (2002) stressed the crucial importance of recursive rules in natural languages, a renewed interest has risen concerning the learnability of recursion. Most studies use the artificial grammar learning (AGL) paradigm (Corballis, 2007; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Perruchet & Rey, 2005). In particular, Fitch and Hauser (2004) proposed that the ability of mastering hierarchical structures was critical to distinguish human and nonhuman primates. They argued that humans could grasp hierarchical structures generated by an A^nB^n grammar (see Figure 1), while tamarins were incapable. Moreover, Bahlmann and Friederici (2006) (henceforth B&F) and Bahlmann, Schubotz and Friederici (2008) carried out an fMRI study to probe into the neural basis of processing long-distance dependencies. Significantly greater blood flow was observed in Broca's area during processing of hierarchical-dependency A^nB^n compared to adjacent-dependency $(AB)^n$.

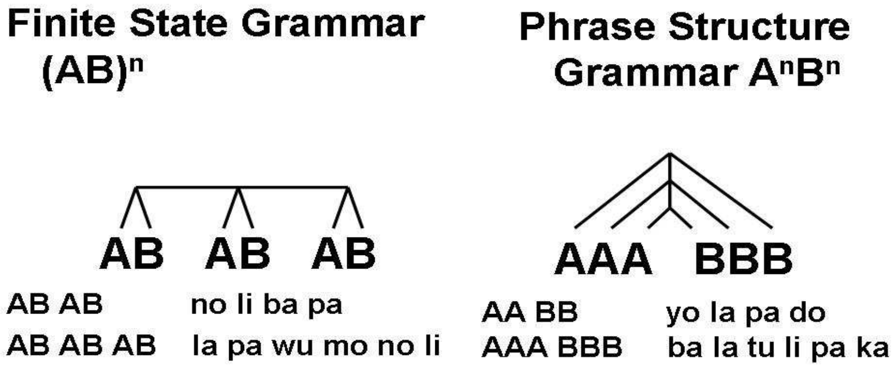


Figure1. Structures of Finite State Grammar $(AB)^n$ and Phrase Structure Grammar A^nB^n used by Fitch and Hauser (2004). Examples of Category A words are: no, ba, la, wu and Category B words are: li, pa, ka, do.

However, as indicated by Perruchet and Rey (2005), the mapping of A-to-B is the essential characteristic of hierarchical center-embedding recursion. At each LoE, this mapping has to be legal according to the grammar[1]. Therefore, Fitch and Hauser (2004), whose grammar did not specify such mapping, could not demonstrate knowledge of center-embeddings in their experiment. The same problem applies for B&F. Though B&F did use a grammar specifying a hierarchical A-B mapping, their test materials were incapable of detecting center-embedded structure learning. When the test materials were controlled, participants failed to learn, as showed by de Vries, Monaghan, Knecht and Zwitserlood (2008), who argued that performance in B&F is based on superficial heuristics, like counting the A's and B's, or repetition-monitoring, instead of learning the center-embedded principle[2].

Previous research has mainly focused on the cognitive learnability of center-embedded structures, rather than on features of the environmental input. Here, we propose two crucial but previously poorly attended environmental factors: One is the organization

---

[1] For instance, $A_1A_2A_3B_3B_2B_1$ is grammatical, whereas $A_1A_2A_3B_1B_2B_3$ is not.

[2] Indeed, in B&F, violations were *replacement violations* (e.g. $A_1A_2A_3B_3A_2B_1$) and *concatenation violations* (e.g.$A_1A_2B_2B_3$). Contrarily, de Vries et al. (2008) tested two other types: *scrambled* (e.g. $A_1A_2A_3B_1B_3B_2$) and *scrambled+repetition* ($A_1A_2A_3B_1B_3B_1$). Their participants could detect the scrambled+repetition violations, but not the scrambled ones.

of the input by stages (*starting small*, henceforth SS) and the second is sufficient exposure to the grammar's basic adjacent-dependencies in the earliest stage of learning. The purpose of the present research is to explore the impact of these two closely-related conditions on learning center-embeddings.

Considering natural language learning, child-directed speech globally satisfies these conditions, as it has, in the earliest stage, short linguistic constituents, simple grammatical constructions, and little syntactical variability (Pine, 1994; Tomasello, 2003). As children grow, child-directed speech develops gradually into more mature speech types (Bellinger, 1980; Garnica, 1977). Hence, the input on which the learning process operates does not come in a random order. Therefore, if we can demonstrate experimentally the facilitation effect of a growing environmental input, and early exposure to zero-level-of-embedding (0-LoE) exemplars, this result might help understanding the role of the environment in complex natural language learning.

The notion of SS was first raised by Elman (1991, 1993). He trained a connectionist network to parse complex structures which contained embedded subordinates. The network succeeded only if provided with a staged input, but not after exposure to the entire input as a whole. Subsequent studies yielded mixed results, though. Some findings are consistent with Elman's effect (Conway, Ellefson, & Christiansen, 2003; Kersten & Earles, 2001; Krueger & Dayan, 2009; Newport, 1988, 1990; Plunkett & Marchman, 1990). However, other research reported no effect of staged-input (Fletcher, Maybery, & Bennett, 2000; Ludden & Gupta, 2000; Rohde & Plaut, 1999).

In the current study, two AGL experiments were carried out using similar materials as B&F and de Vries et al. (2008). In Experiment 1, we compared learning with a staged-input and a random input. Both learning sets contained 0-LoE exemplars. In Experiment 2, 0-LoE learning items were omitted.

Experiment 1

All participants were exposed to the same strings, generated by grammar <u>*G*</u> (Figure 2). In the SS condition, syllable strings were presented progressively according to their LoE.[3] In the random condition, exactly the same set was presented randomly. We hypothesize that the SS group outperforms the random group.

Method

Participants. Twenty-eight students (20 female), from Leiden University participated. All were native Dutch speakers.

Materials and design. There were two sets of syllables, categorized by their vowels. Category A contained -e/-i, i.e. {be, bi, de, di, ge, gi}, whereas Category B contained -o/-u, i.e. {po, pu, to, tu, ko, ku} (see Appendices A and B). Each A-syllable was connected with its counterparts in Category B according to another cue: their consonants, i.e. {be/bi-po/pu}, {de/di-to/tu} and {ge/gi-ko/ku}. Strings were constructed with two, four, or six paired-syllables following the A^nB^n rule. Frequencies of syllable occurrence were controlled for.

---

[3] For the SS group, in the first four blocks, only 0-LoE learning items were presented. The following four blocks displayed 1-LoE items only. In the last four, 2-LoE items were presented. The ordering of strings within one block was counterbalanced over participants.

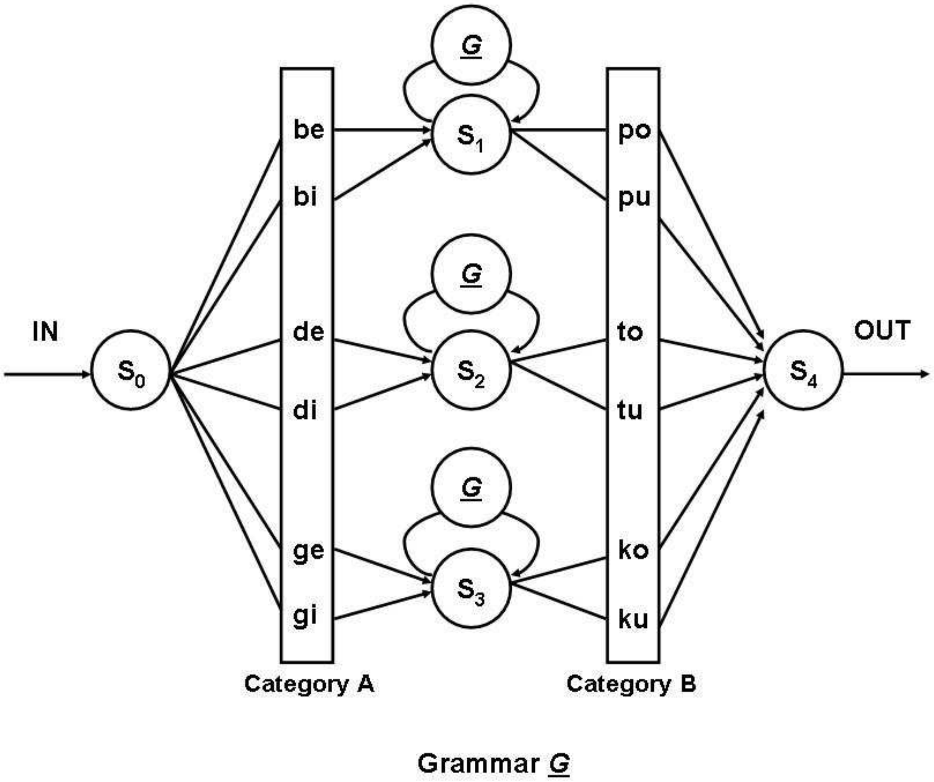


Figure 2. Grammar G, an A^nB^n center-embedded structure. The grammar starts from S_0 and follows one of all possible paths until S_4. "G" in the loops at states S_1, S_2 and S_3 refer to the self-referential rule, indicating that a center-embedded clause can legally be inserted at that specific state. Examples of strings generated by G are: bi pu (0-loE), de ge ko tu (1-loE), be di ge ku to po (2-loE).

The experiment consisted of 12 blocks, with a learning phase and a testing phase each. Twelve strings were presented in each learning phase, and 12 novel strings in each testing phase, of which six were grammatical and six ungrammatical. Both groups were presented the same test strings with 0-, 1-, or 2-LoE. Ungrammatical strings were created by mismatching A-syllables with B-syllables. For two-syllable strings, violations appeared necessarily in the second position ($A_1\underline{\mathbf{B_2}}$); for four-syllable strings, in the fourth position ($A_1A_2B_2\underline{\mathbf{B_3}}$); and for six-syllable strings, in the fifth or sixth position ($A_1A_2A_3B_3\underline{\mathbf{B_4}}B_{1,}$ $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$). For instance, the violation B_4 in $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$ means that the last B mismatches any A in this sequence. In this manner, no adjacent AB violations in the middle of a string could occur, except, necessarily, for two-syllable test strings. Moreover, in contrast to B&F, no repetition of exactly the same syllable appeared in the same sequence, and all test strings had an equal number of A's and B's. As a result, violations could not easily be detected on the basis of surface heuristics or bigram violations.

Procedure. Participants were informed that they would see strings satisfying a sequential rule. Each learning trial started with a fixation cross (500 ms). Then, each syllable was presented separately for 800 ms, with no interval in-between[4].After presentation of 12 strings, a testing phase followed. When the last syllable of each test string disappeared, participants had to indicate "YES" or "NO" depending on whether they believed the string satisfied the rule also underlying the learning strings. Feedback was given (500 ms). For ease of comparison with findings by B&F and de Vries et al. (2008), their explicit procedure was also applied in the current study. The task took 30 minutes approximately.

Results and discussion

A t-test on mean *d'*-values[5] revealed that, overall, the SS group, d'= 1.51(73% correct), highly outperformed the random group, d'= .08 (52% correct), *t* (26) = 3.94, *p* = .001. Only the SS group performed above chance, *t* (13) = 4.21, *p* = .001.

---

[4] With this manipulation, we tried to simulate the situation of natural language processing maximally, in the laboratory environment.

[5] Due to a small response bias favoring positive responses (M = .53, SE = .01, p < .01), d'-values were applied as a measure for sensitivity to grammaticality of the responses.

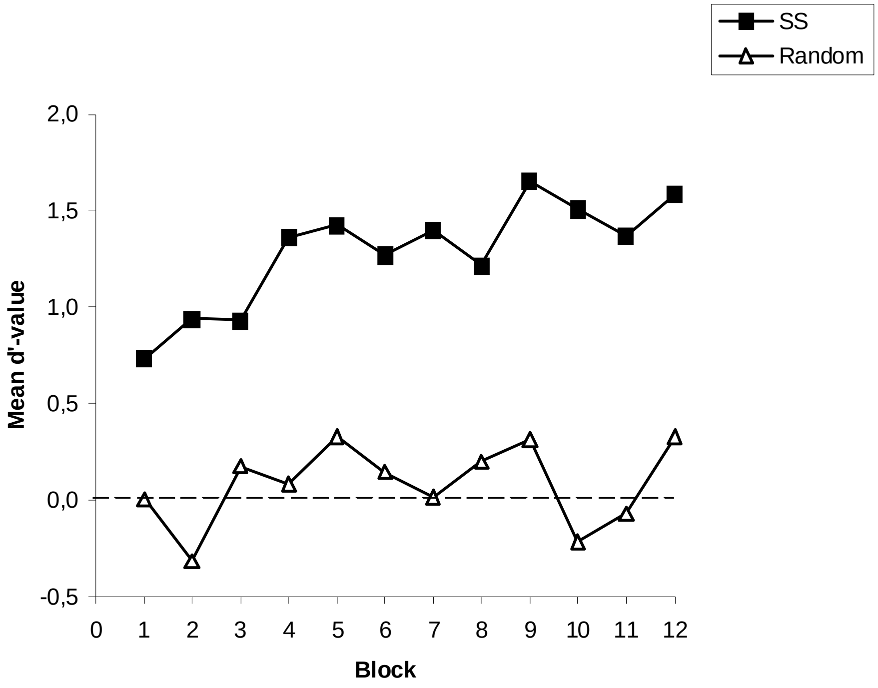
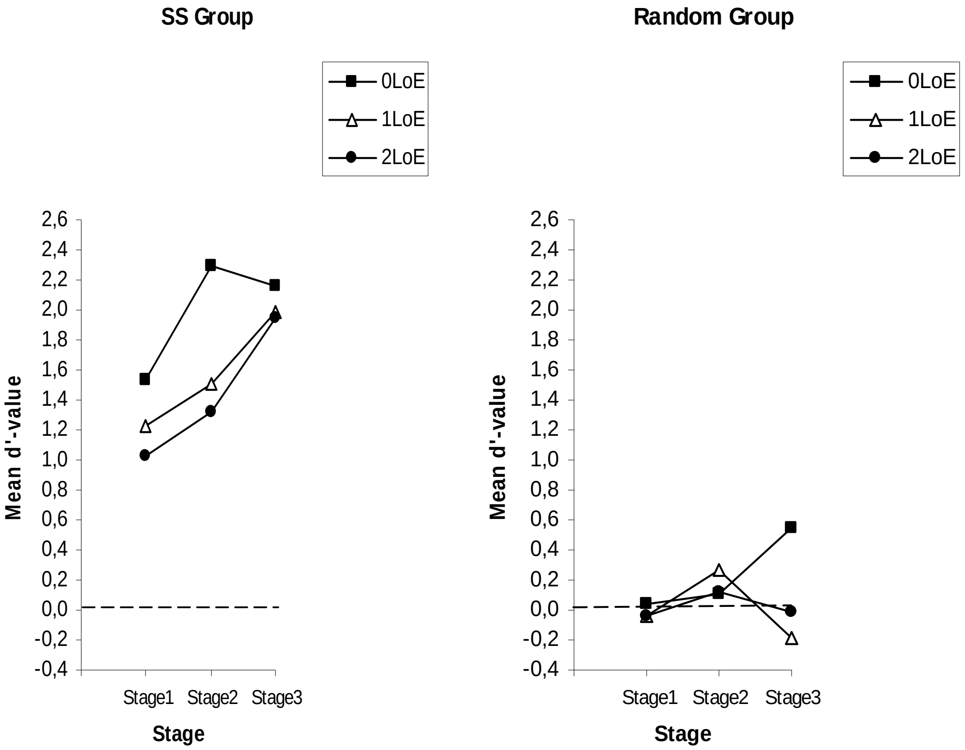


Figure 3. Experiment 1: Mean d'-values for all blocks in both conditions. Points represent mean d'-values per block. The dotted line represents chance level performance (d'= 0).

Moreover, the SS group improved in Block 12, $d'_{12} = 1.59$ (78% correct), compared to Block 1, $d'_1 = .73$ (63% correct), $t\,(13) = 2.59$, $p < .05$. In the random group, however, performance did not improve over time: $d'_1 = .01$ (50% correct), $d'_{12} = .33$ (56% correct), $t\,(13) = -.98$, n.s.. Although in Block 1 the SS group performed slightly better than the random group, this difference was not significant, $t\,(26) = 1.98$, n.s.. However, in the last block, the SS group clearly outscored the random group, $t\,(26) = 2.87$, $p < .01$. In Figure 3, mean d'-values are displayed for all blocks, showing learning in the SS group over time, but not for the random group.

In addition, performance on different types of test items (0-, 1-, and 2-LoE) was compared at several stages of exposure[6]. An ANOVA, with LoE and stage as within-subject factors and condition as between-subject factor showed main effects of LoE, F (2, 52) = 9.00, *p* < .001; of stage, F (2, 52) = 3.92, *p* = .04; and of condition, F (1, 26) = 17.30, *p* < .001. The LoE × Stage × Condition interaction was significant, F (4, 104) = 2.94, *p* = .02, indicating, that performance on various LoE test items developed differently under each condition.



---

[6] Stage 1 consisted of Block 1-4, Stage 2 consisted of Block 5-8, and Stage 3 consisted of Block 9-12 (see Appendix A). Especially for the SS group, Stage 1 comprised 0-LoE learning items only; Stage 2, 1-LoE items only; Stage 3, 2-LoE items only; whereas for the random group, various LoEs were presented in all learning stages.

Figure 4. Experiment 1: Mean d'-values for 0-, 1-, and 2-LoE test items at different stages. Points represent mean d'-values of performance per stage. The dotted line represents chance level performance (d'= 0).

Subsequently, for each group we conducted an ANOVA with LoE and stage as within-subject factors. Under the SS condition, there were main effects of LoE, F (2, 26) = 10.86, $p < .001$, and of stage, F (2, 26) = 3.57, $p < .05$. Performance for 0-LoE items (see Figure 4), d'= 1.89 (77% correct), was significantly better than 1-LoE, d'= 1.45 (72% correct), t (13) = 3.14, $p < .01$ and 2-LoE, d'= 1.29 (70% correct), t (13) = 4.19, $p = .001$, respectively. However, in the random group, chance level performance was observed for all types of test items. There was no effect of LoE, F (2, 26) = 1.31, n.s., neither of stage, F (2, 26) = .87, n.s..

In sum, our findings revealed learning of center-embedded structures in the SS procedure, but not in the random procedure. Moreover, gradual exposure to the staged input, co-occurred with a synchronic improvement in performance. Strikingly, at the end of the first stage, when the SS group had been exposed to 0-LoE only, they performed better (d'= 1.36, 74% correct) than the random group (d'= .08, 52% correct), who did see higher-than-0-LoE learning items, t (26) = 3.42, $p < .005$.

To test further whether performance in the SS group could rely on other strategies, even after careful control for possible confounding surface cues (de Vries et al., 2008) in the test materials, we looked for complex surface calculations that might have underlain detection of particular violations. We subsequently classified these violations according to the surface rule that could possibly have been used to detect them[7]. We then could predict that if knowledge of the center-embedded principle was the basis of response, equal performance on all types of violations, should be found. If, alternately, participants relied on surface cues, different performance may be expected for types of violations detectable with different cues or calculations. In particular, lower performance can be expected as more complex calculations are needed to detect a violation. We found no effect of type of

---

[7] Three types of violations were distinguished: Type I ($A_1A_2A_1B_1B_2B_2$) violation with A's and B's from the same subsets but not equally distributed for the A's as for the B's; Type II ($A_1A_1B_1B_2$, or $A_1A_2A_2B_2B_2B_3$) with a B that could not be paired with any A; Type III ($A_1A_2B_2B_2$, or $A_1A_2A_3B_3B_2B_2$), with one A missing a B from the same subset. Indices here refer to subsets of syllables within A or B category. Each subset consists of two different syllables.

violation on performance, F (2, 26) = .15, n.s.. Participants' performance in the SS group was actually highly similar for all types of violations[8].

A possible surface heuristic that de Vries et al. (2008) paid attention to, is 'monitoring repetitions'. In our materials, no exact repetitions could occur; though repetitions of syllables within the same A or B subcategory could (for example *bebi*- or -*totu* could occur as part of a sequence). However, this type of repetitions was independent of grammaticality of the sequence in our test materials: subset repetitions both occurred in grammatical (e.g., $A_1A_1B_1B_1$) and ungrammatical (e.g., $A_1A_1A_2B_2\underline{\mathbf{B_2}}B_1$) items. Thus, subset repetitions could not be used as a heuristic. Overall, our stimuli and data weaken the possibility that participants used surface rules to perform the grammaticality-judgment task.

Since robust knowledge of 0-LoE exemplars was shown in the SS group only, knowledge of two-syllable sequences might be necessary to grasp the embedding principle. Indeed, primary exposure to adjacent-dependencies was hypothesized to be another crucial factor facilitating learning. We conducted Experiment 2 to verify this hypothesis. We compared again a SS group with a random group, as in Experiment 1, removing all 0-LoE learning items in both conditions.

Experiment 2

Method

Participants. Eighteen students (13 female) from Leiden University participated. None had participated in Experiment 1.

Materials and design. The same materials except 0-LoE learning items were adopted from Experiment 1. Participants were trained with 96 items possessing 1- or 2-LoE (See Appendices C and D). In the learning phase, the SS group was first presented with four blocks of 1-LoE items, and subsequently, with four blocks of 2-LoE items, whereas the random group was presented with the same input randomly.

Procedure. Identical to Experiment 1.

---

[8] Mean accuracy for test items with violation Type I, II, III were .69, .69, and .67 respectively.

Results and discussion

Overall the SS group, d’= .05 (51% correct), did not differ from the random group, d’= .18 (53% correct), *t* (16) = -1.11, n.s.. Both groups performed at chance level. Additionally, for both groups (see Figure 5), performance did not change between the first and the last blocks, d'_1 = -.12 (48% correct), d'_8 = .32 (56% correct), *t* (8) = 1.50, n.s. for the SS group, and d'_1 = .32 (56% correct), d'_8 = .08 (51% correct), *t* (8) = .72, n.s., for the random group. These data indicate that participants could not distinguish grammatical items when no 0-LoE training items presented to them, even in an SS procedure.

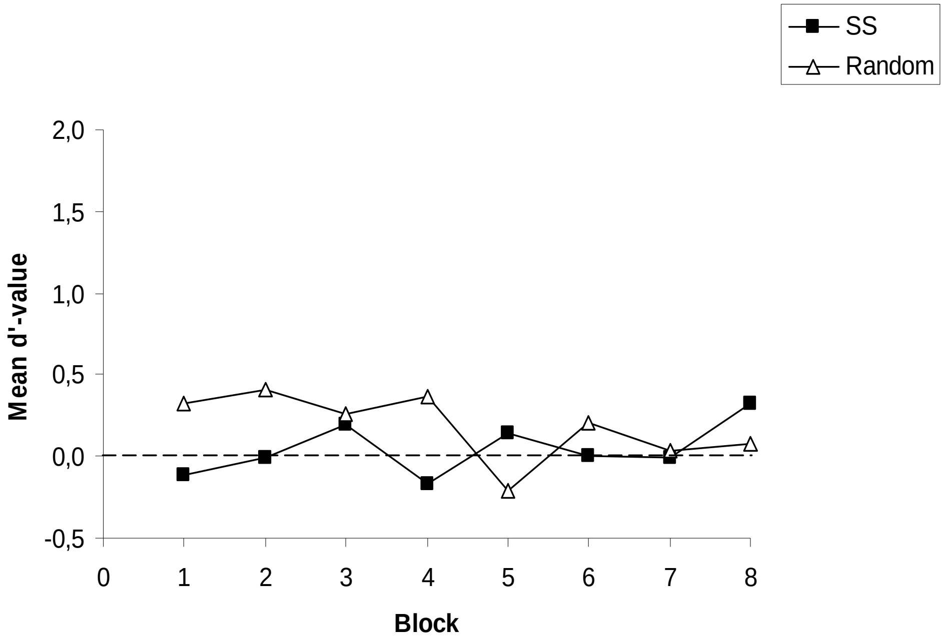


Figure 5. Experiment 2: Mean d’-values for all blocks in both conditions. Points represent mean d’-values of performance per block. The dotted line represents chance level performance (d’= 0).

General discussion

The present research provides insight into two crucial environmental conditions affecting the learnability of a hierarchical center-embedded grammar: first, the effect of an incrementally presented input; second, the importance of exposure to adjacent-structures in the earliest stage of training. Experiment 1 showed that participants performed better on a grammaticality-judgment task after training with an input organized incrementally, according to their LoE. Also, even basic adjacent-dependencies were better learned under SS conditions. The facilitation effect of SS disappeared, as Experiment 2 further revealed, when participants were deprived of exposure to the 0-LoE exemplars. The lack of 0-LoE resulted in an incapability to detect structure, no matter whether the stimuli were presented incrementally or randomly. Clustered exposure to basic adjacencies and a staged-input seem to play crucial roles in learning embedded hierarchical structures.

As previous studies (Christiansen & Dale, 2001; McDonald & Plauche, 1995; Perruchet & Rey, 2005; Poletiek, 2002; Poletiek & Chater, 2006) have suggested, SS may have a better impact when it is assisted by some other cues. The current data indicate that the SS effect can operate if and only if it is combined with sufficient primary exposure to basic adjacent-dependencies of the structure. Especially the striking effect that the SS group outperformed the random group after exposure to 0-LoE only, possibly indicates that once participants were familiarized with the basic associations, they could recognize the associated pairs, even if located in remote positions. Possibly, knowledge of the fundamental adjacent-dependencies serves as a crucial stepping stone in exploring complex hierarchical structures in subsequent stimuli.

The effects of staged-input and early adjacent-dependencies point at the close collaboration between cognition and environment, specifically between an incremental learning mechanism and an incrementally organized input. Thus far, research has mainly focused on the cognitive mechanisms underlying learning complex structures. For instance, a recent fMRI study demonstrated that the activation of the left pars opercularis in processing hierarchical center-embeddings (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006), also occurs during processing of German (Makuuchi, Bahlmann, Anwander, & Friederici, 2009). And several studies with artificial materials have looked at how long-distance-dependencies are processed (Mintz, 2002, 2003; Onnis, Monaghan, Christiansen, & Chater, 2004).

Our study suggests the importance of a good match between cognition and the environment, in facilitating the learning process of hierarchical center-embeddings. This match may also be at work in natural language learning. Although the procedure used in the present lab study (explicit instructions and visual presentation of the stimuli), deviates from the natural language learning context, the facilitating factors we found may be operating in the natural situation as well. Indeed, the natural environment (child directed speech) is incremental and the early learning strategy associative. Some other studies on language learning are in line with this analysis. Gomez & Maye (2005) argue that the ability to associate constituents is important in learning natural syntax, especially since center-embedded recursion is one of its main features. A study on American Sign Language (Newport, 1990) showed that early learners outperformed late learners because the former went through a stage in which they were highly familiarized with the simplest constituents. After that, they could become proficient at combining short constituents into more complex entireties.

Our results also generate new questions. For instance, are hierarchical center-embeddings only learnable after some critical level of prior knowledge on adjacent-dependencies has been obtained? Future work has to find out to what criterion learners have to acquire basic knowledge before increasing input complexity can be processed. Moreover, the frequencies of each LoE-category of training items are also interesting for investigation. A current study in our lab suggests that decreasing numbers of exemplars with increasing complexity are needed for learning the underlying system (Poletiek, Chater, & Van den Bos, submitted). Another question is whether different modalities of exposure would affect performance (Conway & Christiansen, 2005). Finally, it is important to find out the limits of the generalizability of the present and similar data for explaining natural processes. A straightforward question is to what extent the huge complexity of natural grammars might invalidate generalization from the experimental noiseless artificial situation.

In sum, the present study reveals crucial roles for a staged-input and solid primary knowledge of the basic structures, in learning by induction a center-embedded structure. From a more general point of view, our research suggests that the old puzzle of the learnability of hierarchical structures might benefit from a shift of focus on the stimulus environment and its fitness to how human learning works and develops over time.

Footnote

[1] For instance, $A_1A_2A_3B_3B_2B_1$ is grammatical, whereas $A_1A_2A_3B_1B_2B_3$ is not.

[2] Indeed, in B&F, violations were *replacement violations* (e.g. $A_1A_2A_3B_3A_2B_1$) and *concatenation violations* (e.g.$A_1A_2B_2B_3$). Contrarily, de Vries et al. (2008) tested two other types: *scrambled* (e.g. $A_1A_2A_3B_1B_3B_2$) and *scrambled+repetition* ($A_1A_2A_3B_1B_3B_1$). Their participants could detect the scrambled+repetition violations, but not the scrambled ones.

[3] For the SS group, in the first four blocks, only 0-LoE learning items were presented. The following four blocks displayed 1-LoE items only. In the last four, 2-LoE items were presented. The ordering of strings within one block was counterbalanced over participants.

[4] With this manipulation, we tried to simulate the situation of natural language processing maximally in the laboratory environment.

[5] Due to a small response bias favoring positive responses (M = .53, SE = .01, p < .01), d'-values were applied as a measure for sensitivity to grammaticality of the responses.

[6] Stage 1 consisted of Block 1-4, Stage 2 consisted of Block 5-8, and Stage 3 consisted of Block 9-12 (see Appendix A). Especially for the SS group, Stage 1 comprised 0-LoE learning items only; Stage 2, 1-LoE items only; Stage 3, 2-LoE items only; whereas for the random group, various LoEs were presented in all learning stages.

[7] Three types of violations were distinguished: Type I ($A_1A_2A_1B_1B_2B_2$) violation with A's and B's from the same subsets but not equally distributed for the A's as for the B's; Type II ($A_1A_1B_1B_2$, or $A_1A_2A_2B_2B_2B_3$) with a B that could not be paired with any A; Type III ($A_1A_2B_2B_2$, or $A_1A_2A_3B_3B_2B_2$), with one A missing a B from the same subset. Indices here refer to subsets of syllables within A or B category. Each subset consists of two different syllables.

[8] Mean accuracy for test items with violation Type I, II, III were .69, .69, and .67 respectively.

Appendix A

Stimuli in the starting small condition of Experiment 1

	Stage 1			
Phase	Block 1	Block 2	Block 3	Block 4
Learning	bepu	bepu	bepu	bepu
	bepo	bepo	bepo	bepo
	ditu	ditu	ditu	ditu
	dito	dito	dito	dito
	giku	giku	giku	giku
	giko	giko	giko	giko
	bipo	bipo	bipo	bipo
	detu	detu	detu	detu
	bipu	bipu	bipu	bipu
	geko	geko	geko	geko
	deto	deto	deto	deto
	geku	geku	geku	geku
Testing	Grammatical			
	deto	bipu	bepu	giku
	geku	geko	ditu	dito
	dibeputo	debiputu	debeputo	bigekupo
	biditupo	bedetopo	bebipupo	geditoku
	debigekopotu	degebepukotu	gebeditupuku	dibegikuputo
	gidibeputuko	bibeditupopu	gigebipukuko	bigidetukupu
	Ungrammatical			
	bi**ko**	de**ko**	be**tu**	ge**po**
	ge**pu**	ge**to**	gi**to**	de**pu**
	degiko**ku**	digiko**ku**	degeko**po**	dibepo**ko**
	gebepo**pu**	begiku**to**	biditо**ko**	gibipo**to**
	dibegikupu**po**	digebepo**tu**to	begidituko**ku**	dibibepopu**ku**
	bedibipu**ko**po	gedibiputu**po**	digideto**pu**tu	gigeditu**pu**ko

Appendix A (continued)

	Stage 2			
Phase	Block 5	Block 6	Block 7	Block 8
Learning	dedituto	debeputu	debepoto	debipotu
	degikoto	degekutu	degekotu	degikuto
	dibiputu	didetuto	didetotu	dibepotu
	digikuto	digikotu	dibiputo	digekuto
	beditupo	beditopo	bedetupo	bedetopu
	begekupo	begekopu	begikopu	bebipopu
	bidetopo	biditopu	bibepupo	bidetupo
	bigekupu	bigekopu	bigikupu	bigikopo
	gedetuku	gedetoko	gedituku	gebepuko
	gegikuko	gebepoku	gebipuko	gegekoku
	giditoku	gidetoku	gidetoko	gidituku
	gigekuko	gibepuko	gibipuku	gibipuko
Testing	Grammatical			
	dito	bipo	bipu	deto
	bepo	geko	detu	geku
	gegikoku	digekutu	gebipuku	degikutu
	debipoto	bigikopu	gidetuko	dibeputu
	gibegekupoko	bidigikotopu	degeditokotu	gedibipotuko
	dibedetupoto	gedegikutuko	begibipokupu	gidebepotuko
	Ungrammatical			
	be**ko**	gi**tu**	bi**tu**	di**po**
	de**ku**	de**po**	gi**pu**	be**to**
	digeko**po**	biditu**ku**	debepo**ku**	degeku**pu**
	giditu**pu**	gebipo**to**	begiku**ko**	gebepu**tu**
	bigidetoko**tu**	begedetu**pu**po	gegidetuko**to**	bibegekopo**ku**
	bigeditu**to**po	gibeditopu**tu**	gedebipu**po**ko	gedibipu**ko**ku

Appendix A (continued)

	Stage 3			
Phase	Block 9	Block 10	Block 11	Block 12
Learning	dedibepotuto	dedigikututo	debeditoputu	debegekoputu
	degigekokutu	debiditupoto	degebipukotu	degibepukuto
	dibibepupoto	didebepototu	dibegekoputo	dibidetopotu
	digebipukotu	digigekokutu	dibigikuputo	digedetukoto
	bebegekupopu	bebiditupopu	bedegekotupu	bedidetutopo
	begeditokupo	begidetokopo	bebigikupupo	begebipokopu
	bidibeputupo	bibedetopopu	bidegikutopu	bidibepotopu
	bigeditokopo	bigedetukupo	bigigekukopo	bibiditupopu
	gedegikotoku	gedibeputuku	gedigikutoko	gedebipotuko
	gebebipopuku	gebebipupoko	gegebepokuko	gegibipokoku
	gibedetupoko	gidibipotoko	gidegekotuku	gidebipotoko
	gibebipupoko	gibidetopuku	gigeditokoku	gigidetukoku
Testing	Grammatical			
	bepu	ditu	detu	bipo
	giko	giko	bepo	giku
	begekopo	gibepuku	bedetupu	gedetoku
	gebipoko	digikutu	bidetopu	gibepoku
	bedidetutopu	gegibepukuko	bedigekotopu	debegekoputo
	bididetotupu	debibepoputu	bigiditukopu	digebepukutu
	Ungrammatical			
	ge**tu**	bi**to**	di**ko**	di**ku**
	di**pu**	be**ku**	bi**ku**	gi**po**
	geditu**po**	dibipo**pu**	beditu**to**	bibepo**tu**
	begiko**ku**	begeku**tu**	gibepo**tu**	gedetu**pu**
	bedegekotu**to**	bidibepo**ko**pu	gididieto**po**ko	debigeku**to**tu
	dibegeku**ko**tu	degibipuko**po**	gedegikuto**tu**	debiditupu**ku**

Appendix B

Stimuli in the random condition of Experiment 1

	Stage 1			
Phase	Block 1	Block 2	Block 3	Block 4
Learning	bebigikupupo	dibidetopotu	bebiditupopu	dibepotu
	bigikopo	gidetoku	ditu	bipo
	gebipuko	dibibepupoto	gibipuko	bigekupu
	gedituku	bepu	deto	bigeditokopo
	beditopo	degikuto	bebipopu	giku
	bigigekukopo	bipu	begebipokopu	dedigikututo
	degekutu	dibegekoputo	gibedetupoko	didetotu
	debiditupoto	bidibeputupo	digikuto	degigekokutu
	bidegikutopu	giko	debepoto	bedetupo
	debeditoputu	dibigikuputo	bigikupu	ditu
	geko	gedebipotuko	gegebepokuko	begeditokupo
	degikoto	detu	gigekuko	bipu
	Grammatical			
Testing	deto	bipu	bepu	giku
	geku	geko	ditu	dito
	dibeputo	debiputu	debeputo	bigekupo
	biditupo	bedetopo	bebipupo	geditoku
	debigekopotu	degebepukotu	gebeditupuku	dibegikuputo
	gidibeputuko	bibeditupopu	gigebipukuko	bigidetukupu
	Ungrammatical			
	bi**ko**	de**ko**	be**tu**	ge**po**
	ge**pu**	ge**to**	gi**to**	de**pu**
	degiko**ku**	digiko**ku**	degeko**po**	dibepo**ko**
	gebepo**pu**	begiku**to**	bidito**ko**	gibipo**to**
	dibegikupu**po**	digebepo**tu**to	begiditukoku	dibibepopu**ku**
	bedibipu**ko**po	gedibiputu**po**	digideto**pu**tu	gigeditu**pu**ko

Appendix B (continued)

	Stage 2			
Phase	Block 5	Block 6	Block 7	Block 8
Learning	digekuto	gigeditokoku	bedidetutopo	begikopu
	gibepuko	giko	digikotu	detu
	begekupo	geko	bepu	bidetopo
	gedetuku	bidetupo	gibidetopuku	degekotu
	gidibipotoko	detu	geku	bigedetukupo
	geku	bibepupo	bipo	bepo
	degibepukuto	gidituku	gegikuko	debeputu
	bepo	geko	bibedetopopu	geku
	bibiditupopu	giko	gebepuko	bedetopu
	digebipukotu	dedituto	debipotu	giku
	bipu	bipu	deto	bepo
	geku	gidegekotuku	gedegikotoku	gebepoku
Testing	Grammatical			
	dito	bipo	bipu	deto
	bepo	geko	detu	geku
	gegikoku	digekutu	gebipuku	degikutu
	debipoto	bigikopu	gidetuko	dibeputu
	gibegekupoko	bidigikotopu	degeditokotu	gedibipotuko
	dibedetupoto	gedegikutuko	begibipokupu	gidebepotuko
	Ungrammatical			
	be**ko**	gi**tu**	bi**tu**	di**po**
	de**ku**	de**po**	gi**pu**	be**to**
	digeko**po**	biditu**ku**	debepo**ku**	degeku**pu**
	giditu**pu**	gebipo**to**	begiku**ko**	gebepu**tu**
	bigidetoko**tu**	begedetu**pu**po	gegidetuko**to**	bibegekopo**ku**
	bigeditu**to**po	gibeditopu**tu**	gedebipu**po**ko	gedibipu**ko**ku

Appendix B (continued)

	Stage 3			
Phase	Block 9	Block 10	Block 11	Block 12
Learning	gegibipokoku	gebebipupoko	didetuto	dedibepotuto
	beditupo	gedetoko	didebepototu	digigekokutu
	bepu	bedegekotupu	detu	digedetukoto
	debegekoputu	geko	giku	giko
	biditopu	bepo	gibebipupoko	gedibeputuku
	dito	gigidetukoku	gegekoku	dito
	bebegekupopu	dibiputo	dito	gebebipopuku
	dibiputu	bipo	gedigikutoko	giditoku
	bipo	begidetokopo	deto	gidetoko
	bigekopu	deto	begekopu	gidebipotoko
	giku	bidibepotopu	ditu	gibipuku
	ditu	dito	degebipukotu	bepu
Testing	Grammatical			
	bepu	ditu	detu	bipo
	giko	giko	bepo	giku
	begekopo	gibepuku	bedetupu	gedetoku
	gebipoko	digikutu	bidetopu	gibepoku
	bedidetutopu	gegibepukuko	bedigekotopu	debegekoputo
	bididetotupu	debibepoputu	bigiditukopu	digebepukutu
	Ungrammatical			
	ge**tu**	bi**to**	di**ko**	di**ku**
	di**pu**	be**ku**	bi**ku**	gi**po**
	geditu**po**	dibipo**pu**	beditu**to**	bibepo**tu**
	begiko**ku**	begeku**tu**	gibepo**tu**	gedetu**pu**
	bedegekotu**to**	bidibepo**ko**pu	gidideto**po**ko	debigeku**to**tu
	dibegeku**ko**tu	degibipuko**po**	gedegikuto**tu**	debiditupu**ku**

Appendix C

Stimuli in the starting small condition of Experiment 2

Phase	Block 1	Block 2	Block 3	Block 4
Learning	dedituto	debeputu	debepoto	debipotu
	degikoto	degekutu	degekotu	degikuto
	dibiputu	didetuto	didetotu	dibepotu
	digikuto	digikotu	dibiputo	digekuto
	beditupo	beditopo	bedetupo	bedetopu
	begekupo	begekopu	begikopu	bebipopu
	bidetopo	biditopu	bibepupo	bidetupo
	bigekupu	bigekopu	bigikupu	bigikopo
	gedetuku	gedetoko	gedituku	gebepuko
	gegikuko	gebepoku	gebipuko	gegekoku
	giditoku	gidetoku	gidetoko	gidituku
	gigekuko	gibepuko	gibipuku	gibipuko
Testing	Grammatical			
	dito	bipo	bipu	deto
	bepo	geko	detu	geku
	gegikoku	digekutu	gebipuku	degikutu
	debipoto	bigikopu	gidetuko	dibeputu
	gibegekupoko	bidigikotopu	degeditokotu	gedibipotuko
	dibedetupoto	gedegikutuko	begibipokupu	gidebepotuko
	Ungrammatical			
	be**ko**	gi**tu**	bi**tu**	di**po**
	de**ku**	de**po**	gi**pu**	be**to**
	digeko**po**	biditu**ku**	debepo**ku**	degeku**pu**
	giditu**pu**	gebipo**to**	begiku**ko**	gebepu**tu**
	bigidetoko**tu**	begedetu**pu**po	gegidetuko**to**	bibegekopo**ku**
	bigeditu**to**po	gibeditopu**tu**	gedebipu**po**ko	gedibipu**ko**ku

Appendix C (continued)

Phase	Block 5	Block 6	Block 7	Block 8
Learning	dedibepotuto	dedigikututo	debeditoputu	debegekoputu
	degigekokutu	debiditupoto	degebipukotu	degibepukuto
	dibibepupoto	didebepototu	dibegekoputo	dibidetopotu
	digebipukotu	digigekokutu	dibigikuputo	digedetukoto
	bebegekupopu	bebiditupopu	bedegekotupu	bedidetutopo
	begeditokupo	begidetokopo	bebigikupupo	begebipokopu
	bidibeputupo	bibedetopopu	bidegikutopu	bidibepotopu
	bigeditokopo	bigedetukupo	bigigekukopo	bibiditupopu
	gedegikotoku	gedibeputuku	gedigikutoko	gedebipotuko
	gebebipopuku	gebebipupoko	gegebepokuko	gegibipokoku
	gibedetupoko	gidibipotoko	gidegekotuku	gidebipotoko
	gibebipupoko	gibidetopuku	gigeditokoku	gigidetukoku
Testing		Grammatical		
	bepu	ditu	detu	bipo
	giko	giko	bepo	giku
	begekopo	gibepuku	bedetupu	gedetoku
	gebipoko	digikutu	bidetopu	gibepoku
	bedidetutopu	gegibepukuko	bedigekotopu	debegekoputo
	bididetotupu	debibepoputu	bigiditukopu	digebepukutu
		Ungrammatical		
	ge**tu**	bi**to**	di**ko**	di**ku**
	di**pu**	be**ku**	bi**ku**	gi**po**
	geditu**po**	dibipo**pu**	bedito**to**	bibepo**tu**
	begiko**ku**	begeku**tu**	gibepo**tu**	gedetu**pu**
	bedegekotu**to**	bidibepo**ko**pu	gidideto**po**ko	debigeku**to**tu
	dibegeku**ko**tu	degibipuko**po**	gedegikuto**tu**	debiditupu**ku**

Appendix D

Stimuli in the random condition of Experiment 2

Phase	Block 1	Block 2	Block 3	Block 4
Learning	debipotu	gegekoku	digigekokutu	dibegekoputo
	begeditokupo	digebipukotu	gidebipotoko	dibigikuputo
	gidituku	bedegekotupu	gidibipotoko	gegikuko
	digikotu	gedigikutoko	bidibeputupo	bigekopu
	gedetuku	bibedetopopu	debeditoputu	digikuto
	bidibepotopu	biditopu	gedegikotoku	degibepukuto
	gidegekotuku	gibepuko	dibidetopotu	bebegekupopu
	gebebipupoko	bibepupo	debegekoputu	gedituku
	degekotu	bibiditupopu	gibedetupoko	dedituto
	bidetopo	bigeditokopo	gidetoku	debepoto
	gibebipupoko	dibiputu	bigikupu	gigeditokoku
	digedetukoto	dedibepotuto	gedibeputuku	gidetoko
Testing	Grammatical			
	dito	bipo	bipu	deto
	bepo	geko	detu	geku
	gegikoku	digekutu	gebipuku	degikutu
	debipoto	bigikopu	gidetuko	dibeputu
	gibegekupoko	bidigikotopu	degeditokotu	gedibipotuko
	dibedetupoto	gedegikutuko	begibipokupu	gidebepotuko
	Ungrammatical			
	be**ko**	gi**tu**	bi**tu**	di**po**
	de**ku**	de**po**	gi**pu**	be**to**
	digeko**po**	biditu**ku**	debepo**ku**	degeku**pu**
	giditu**pu**	gebipo**to**	begiku**ko**	gebepu**tu**
	bigidetoko**tu**	begedetu**pu**po	gegidetuko**to**	bibegekopo**ku**
	bigeditu**to**po	gibeditopu**tu**	gedebipu**po**ko	gedibipu**ko**ku

Appendix D (continued)

Phase	Block 5	Block 6	Block 7	Block 8
Learning	gibipuku	dedigikututo	beditupo	gebipuko
	degebipukotu	dibibepupoto	gegebepokuko	begekopu
	bigigekukopo	bigedetukupo	dibiputo	bebigikupupo
	digekuto	degikoto	degikuto	bebiditupopu
	bigikopo	gedetoko	bedetupo	gegibipokoku
	begidetokopo	didetotu	gigidetukoku	degigekokutu
	degekutu	dibepotu	debeputu	begekupo
	gibidetopuku	bidetupo	begebipokopu	debiditupoto
	gebepoku	gebebipopuku	gebepuko	begikopu
	bigekupu	bedidetutopo	didebepototu	gigekuko
	bidegikutopu	gibipuko	bebipopu	bedetopu
	gedebipotuko	giditoku	beditopo	didetuto
Testing	Grammatical			
	bepu	ditu	detu	bipo
	giko	giko	bepo	giku
	begekopo	gibepuku	bedetupu	gedetoku
	gebipoko	digikutu	bidetopu	gibepoku
	bedidetutopu	gegibepukuko	bedigekotopu	debegekoputo
	bididetotupu	debibepoputu	bigiditukopu	digebepukutu
	Ungrammatical			
	ge**tu**	bi**to**	di**ko**	di**ku**
	di**pu**	be**ku**	bi**ku**	gi**po**
	geditu**po**	dibipo**pu**	beditu**to**	bibepo**tu**
	begiko**ku**	begeku**tu**	gibepo**tu**	gedetu**pu**
	bedegekotu**to**	bidibepo**ko**pu	gidideto**po**ko	debigeku**to**tu
	dibegeku**ko**tu	degibipuko**po**	gedegikuto**tu**	debiditupu**ku**

Chapter 3

How “Small” Is “Starting Small” for Learning Hierarchical Center-embedded Structures?

This chapter is based on: Lai, J. & Poletiek, F. (in press). How “small” is “staring small” for learning hierarchical center-embedded structures? *Journal of cognitive psychology*.

Abstract

Hierarchical center-embedded structures pose a large difficulty for language learners due to their complexity. A recent artificial grammar learning study (Lai & Poletiek, 2011) demonstrated a starting-small (SS) effect, i.e. staged-input and sufficient exposure to 0-level-of-embedding exemplars were the critical conditions in learning A^nB^n structures. The current study aims to test: 1) a more sophisticated type of SS (a gradually rather than discretely growing input); 2) the frequency distribution of the input. The results indicate that SS optimally works under other conditional cues, such as a skewed frequency distribution with simple stimuli being more numerous than complex ones.

To the great interest of linguists and psychologists, children display an amazing ability in extracting rules from language and producing new sentences which obey the rules. Especially, how humans process complex recursive center-embedded structures with long-distance dependencies, such as “the rat that the dog that the man walked chased ran” is still poorly explained (Corballis, 2007). Moreover, the learnability of this type of structures has become a major issue in language learning research, since recursion has been proposed to be the crucial feature of the human language faculty (Hauser, Chomsky, & Fitch, 2002). One implication of this position is that such structures cannot be learned from environmental stimuli only and by using general cognitive learning mechanisms. The environment contains too little information to induce rules of recursive complexity, and general learning mechanisms are linear, whilst the system to be learned is hierarchical. This point of view is in line with the *poverty of stimulus* hypothesis (Chomsky, 1980; Perfors, Tenenbaum, & Regier, 2011), which proposes that the accessible data are so impoverished that children are unable to induce and generalize structures from these data to acquire full knowledge of the language system. Therefore, natural language grammar learning must be assisted by an inborn device, according to this reasoning. Indeed, the intrinsic properties of recursion, especially center-embeddings and the corresponding long-distance dependencies, actually pose difficulties for language learners, both in perception and production (Christiansen & Chater, 1999; Gibson, 1998).

A growing body of work attempts to probe into the fundamental cognitive mechanism of learning hierarchical center-embedded structures (Friederici, 2004; Hochmann, Azadpour, & Mehler, 2008). Except for the *starting small* (henceforth SS) effect (Elman, 1991), however, the influence of facilitative factors in learning center-embedding has hardly been investigated experimentally with artificial grammar in the laboratory environment. Elman (1993) trained a simple recurrent network in a word prediction task to learn the underlying rule of the given grammar. The network first failed to learn when it was exposed to the whole set of input, but then succeeded when being presented with an incremental input. This study showed an advantage of limitation of the input resources. Elman (1993) pioneered the concept that a simple recurrent network could learn sentences containing multiple hierarchical embeddings if it was first confronted with simple structure before stepping further into more complex compound sentences with sub-clauses. In line with Elman, Lai and Poletiek (2011) also observed this SS effect by manipulating the organization of the input. In addition, we found that early presentation of a cluster of simple exemplars without embeddings was a prerequisite for learning the center embedded structure from the embedded sentences presented later on.

A similar hypothesis to the SS effect was proposed by Newport (1990), who showed that early learners of American Sign Language were able to achieve higher competence because they started processing limited individual parts first; whereas late learners who began with complete signs as wholes had more difficulties. However, in contrast to Elman who focused on the structure of staged-input, Newport emphasized the internal limitation of cognitive capacity, which actually aided children in successful learning, reducing the units to short sequences in the earliest stage. Empirical evidence also came from Kersten and Earles (2001), who found that adults learned a miniature artificial language better, when they were exposed to an initial training of small constituents, instead of complete sentences. A more general argument was made by Kareev, Lieberman and Lev (1997), who proposed that due to limitation of working memory capacity, people concentrated on small samples of information, which enlarged the possibility of early detection of correlations in the sample (Hertwig & Todd, 2003).

Some other studies have obtained results contradicting the SS facilitation, however. In two simulation studies, Rohde and Plaut (1999) found no facilitation by SS, but instead an advantage of “starting big” in the presence of semantic constraints. With a third simulation, they excluded the possibility that the constrained memory of the network facilitated learning. Therefore, Rohde and Plaut (1999, 2003) suggested that neither staged-input nor restriction of memory was a necessary prerequisite for learning complex statistical regulations. Looking further into the role of cognitive capacity, Ludden and Gupta (2000) stated that the more cognitive resources were provided, the better performance that learners could achieve. Also, older children, and intellectually gifted children showed better learning in an implicit learning task, compared to younger, or intellectually delayed children with limited cognitive capacity (Fletcher, Maybery, & Bennett, 2000). As a final example of SS tests, Conway, Ellefson and Christiansen (2003) found that participants were assisted in learning both nested and right-branching recursion by the SS input only under the visual modality, but not the auditory modality. Hence, not all learners, and not under all conditions do learners benefit from a growing input.

The purpose of the present research is to explore under what additional conditions of the environmental input, SS does facilitate learning. In particular, we suggest that the frequency distribution of the input exemplars may moderate the influence of the SS effect: A starting small ordering might be most helpful if the simplest exemplars of the grammar not only occur in the earliest stage, but also in higher frequency than the more complex exemplars. In the present work, we aim to test the effects of different types of SS ordering, frequency distribution, and their combination. As a variation of the traditional SS organization of the input, we let the input grow smoothly, by inserting more complex stimuli gradually, rather than in clusters. By manipulating the frequency distribution, we further evaluate how much preliminary exposure to the simple structures with zero-level-of-embedding (0-LoE) is needed to enhance complex structure learning.

Frequency distribution of the input has been suggested to play a role in inducing structure from that input. For instance, in a categorization task, adults showed better performance in speech perception by the use of frequency distribution cues of acoustic-phonetic information (Clayards, Tanenhaus, Aslin, & Jacobs, 2008). Moreover, previous studies with children indicated that a skewed distribution facilitated learning new constructions. For instance, Casenhiser and Goldberg (2005) showed that the more frequent a particular single verb was, the better that children learned and generalized the mapping between its form and meaning. Similarly, Kidd, Lieven, and Tomasello (2010) found that high lexical frequency largely boosted children’s learning on sentences which contained verbs in high frequency.

In addition, frequency distribution has also been shown to enable adults to learn non-recursive grammatical features. Poletiek and Chater (2006) presented two groups of participants with the same unique exemplars of an artificial finite state grammar, but in two different frequency distributions: One followed the distribution of a natural random output of the grammar, i.e. short and simple exemplars were presented more frequently than long ones, as they were also more frequently repeated in a random output sample; the other distribution was even, i.e. each unique exemplar was presented an equal number of times, disregarding its length. The group exposed to the “natural” random output of the grammar performed better on a grammaticality-judgment task than the group exposed to the equally distributed input.

Poletiek and Chater (2006), however, used a non-recursive finite state grammar. The role of frequency distribution as a cue for inducing structure might also apply to complex recursive grammar learning. Moreover, if the skewness of the input effectively influences the learnability of complex structures, this might explain the twofold findings by Lai and Poletiek (2011): Center-embedded structure learning requires a combination of both a SS regimen and early exposure to a relatively large cluster of short sequences *without* embeddings. Indeed, successful grammar induction might involve two separate and consecutive learning procedures, requiring 1) early massive exposure to short and simple 0-LoE sentences for grasping the basic pattern of language; 2) after that, a smaller number of 1- and 2-LoE items suffice for learning the recursive operation. In such a two staged learning process, the familiarity of 0-LoE assists human parsers in detecting related elements in more complex items with embedded clauses, showing up in the stimulus set later on. Furthermore, we hypothesize that as exposure to 0-LoE items is more extensive, the detection of pairs in later materials is easier.

Hence we suggest that learners would be helped in grasping a structure by being exposed to more frequent occurrences of simple items it generates, and less frequent complex ones (Casenhiser & Goldberg, 2005; Clayards et al., 2008; Poletiek & Chater, 2006). Notice that this skewed distribution resembles the Zipfian distribution reflected in natural languages (Kurumada, Meylan, & Frank, 2011), in which short and simple constructions occur extremely more often than long and complex occurrences of the grammar.

In the present experiments, we manipulate frequency distribution (equal versus unequal) and ordering (in three ways: the clustered SS set up as in Lai & Poletiek, 2011; a gradual SS regimen, i.e. inserting gradually more complex items over time. This gradual SS condition might be more similar to natural learning situations with increasingly complex input; and a random ordering). These manipulations make it possible to evaluate: first, two different types of SS procedures; second, the effect of early exposure to a cluster of simple sentences; and third, the overall effect of frequency distribution of the input.

Experiment 1

In Experiment 1, we compare three input orderings: first a discrete SS regimen with items clustered by the number of LoE; second, an incremental SS ordering; third, a random ordering. Participants were randomly assigned to one of the groups.

Method

Participants. Forty-five students from Leiden University participated. All were native Dutch speakers.

Materials and design. Grammar *G* with an A^nB^n center-embedded structure in Lai and Poletiek (2011) was used. Yet, a novel set of 120 learning strings was generated (Appendix A). Strings were composed of syllables from Category A, i.e. {be, bi, de, di, ge, gi}, and Category B, i.e. {po, pu, to, tu, ko, ku}. Pairs were specified by the consonants, i.e. {be/bi-po/pu}, {de/di-to/tu} and {ge/gi-ko/ku}. Strings with three different lengths (two, four, or six paired-syllables) were applied. Syllable occurrences were balanced in frequencies. The same number of test items was also produced, half grammatical and half ungrammatical (Appendix C). The violations were constructed by mismatching the specific pairing between A- and B-syllables (e.g. $A_1\underline{\mathbf{B_3}}$; $A_1A_2B_2\underline{\mathbf{B_3}}$; $A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$, or $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$). [1] Violations were not allowed in the middle AB position (except for 0-LoE, in which they were the only possible violation), since an ungrammatical A**B** bigram would be too salient and be easily recognized just by monitoring the superficial characteristics of test items.[2]

Each group was presented with 40 learning items for each LoE. In total, there were 12 blocks, with a learning phase (10 items) and a testing phase (10 items) in each. In the

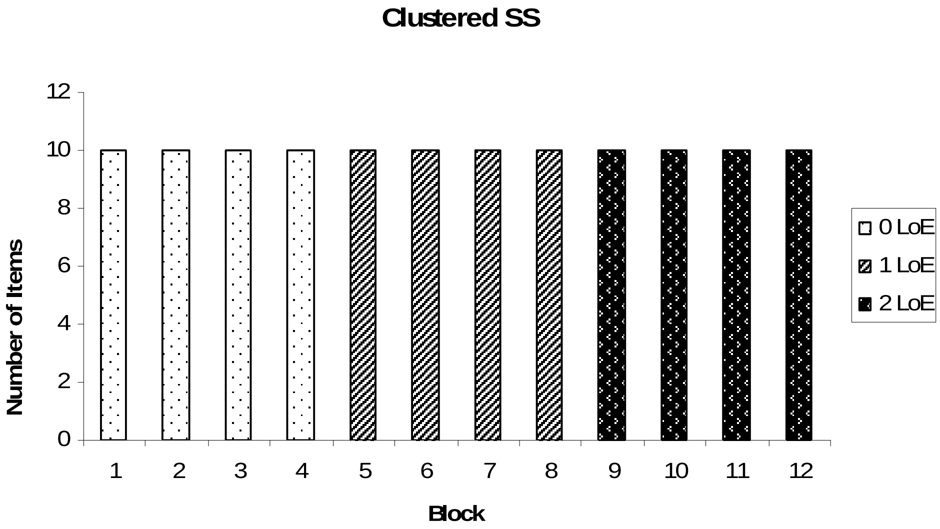
---

[1] In order to avoid easy detection with the hint of surface heuristics, no repetition of exactly the same syllable was allowed in the same string. In the test string, the number of A's and B's is equal.

[2] This criterion results in: 1) for 1-LoE, the violations would always appear at the last position (e.g. $A_1A_2B_2\underline{\mathbf{B_3}}$); 2) for 2-LoE, we equally divided ungrammatical items into two types of violations: one type with violations at the last position (e.g. $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$), and the other with violations at the second-to-the-last position (e.g. $A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$).

learning phase, the ordering of items was manipulated (Figure 1): For the clustered SS group, participants would first see 0-LoE learning items only in the first four blocks, then only 1-LoE in Block 5-8 and 2-LoE in Block 9-12. For the incremental SS group, participants would first see only 0-LoE in the first block; From Block 2 on, a few 1-LoE items were introduced gradually and in Block 6, 2-LoE items were introduced. As more complex items were displayed, the number of lower level ones decreased. For the random group, the same material was presented in a randomized order.

All groups were tested with the same items.



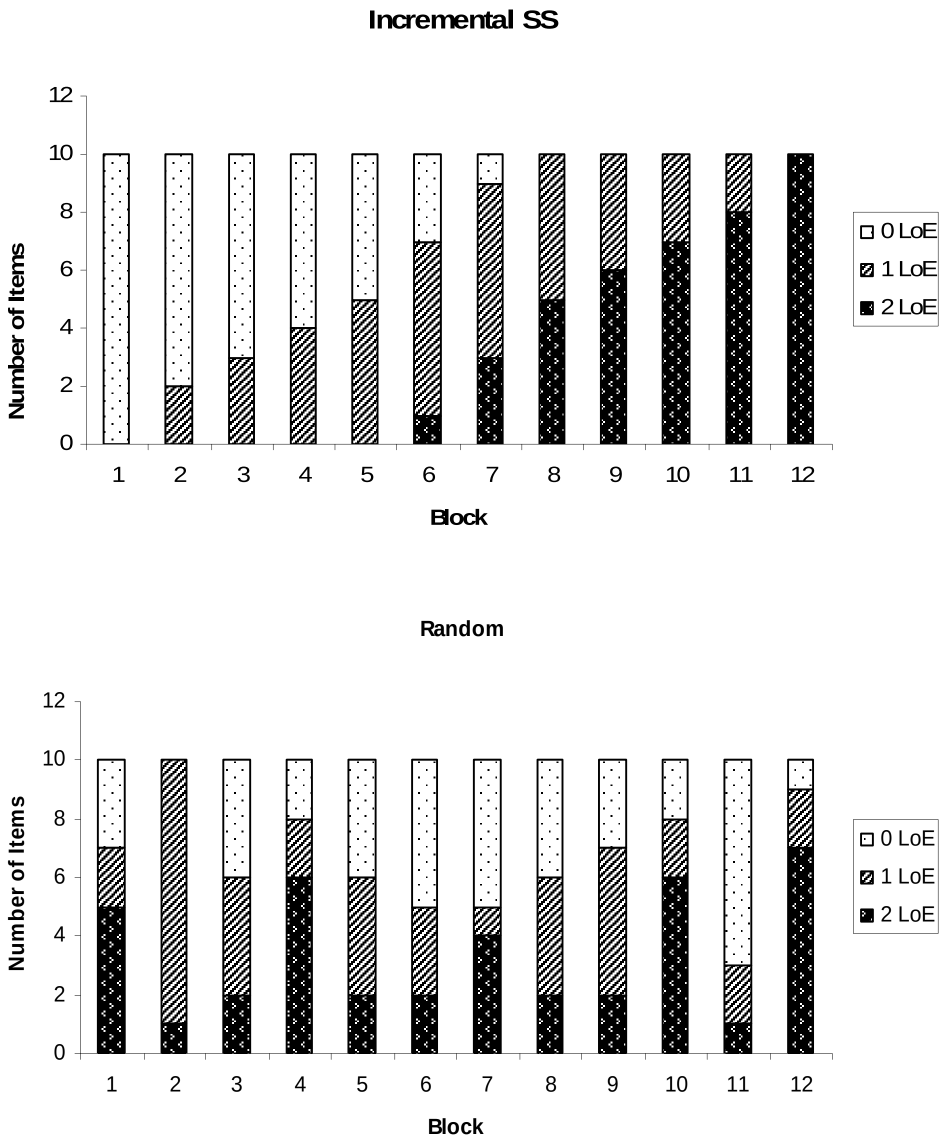


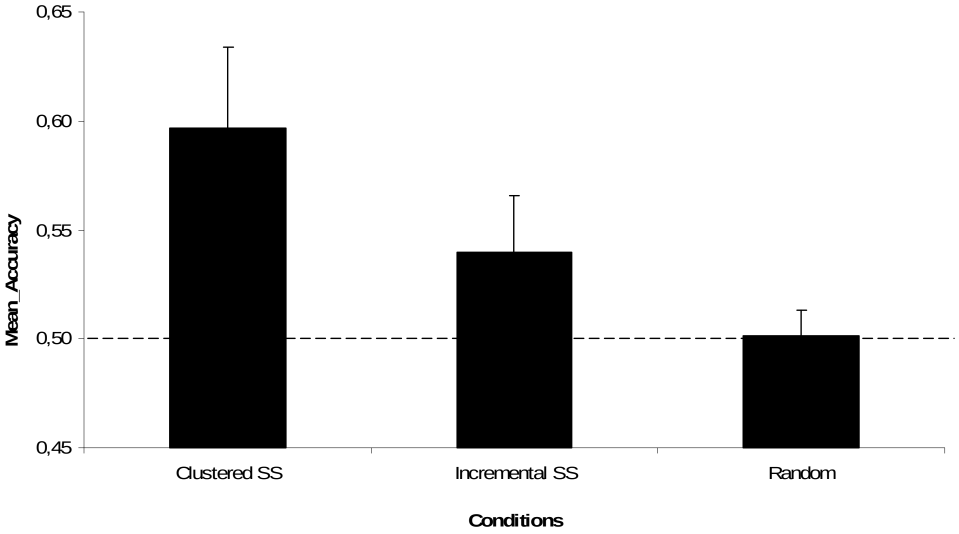
Figure 1. Experiment 1. The ordering of exemplars with 0-, 1-, and 2-LoE in the input under the clustered SS, the incremental SS, and the random condition.

Procedure. In the learning phase, participants were instructed that the syllable strings presented were governed by an underlying rule. In each trial, after a fixation cross (500 ms), a learning item was presented syllable-by-syllable visually (800 ms per syllable, with no interval in-between). Participants would see 10 learning items consecutively. Next, 10 novel items were presented in the same way in the test phase, for which grammaticality-judgments were required. Feedback was given (500 ms).

Results and discussion

We compared performance over the entire set of 12 blocks for different groups. An ANOVA showed a main effect of condition, F (2, 42) = 3.23, $p < .05$, $\eta_p^2 = .13$. As displayed by Figure 2.a, only the clustered SS group (M = .60, SE = .04) performed significantly above chance, t (14) = 2.64, $p < .05$, $r = .58$. T-tests showed that the clustered SS group performed significantly better than the random group (M = .50, SE = .01), t (28) = 2.48, $p < .05$, $r = .42$; yet, there was no significant difference between the clustered SS group and the incremental SS group (M = .54, SE = .03), t (28) = 1.26, n.s.; nor between the incremental SS group and the random group, t (28) = 1.40, n.s.

a)



b)

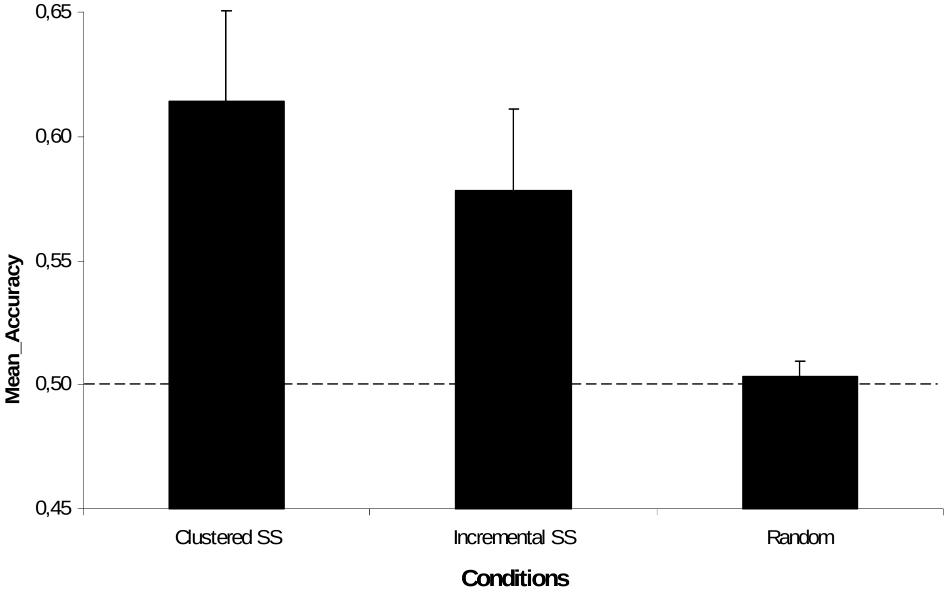
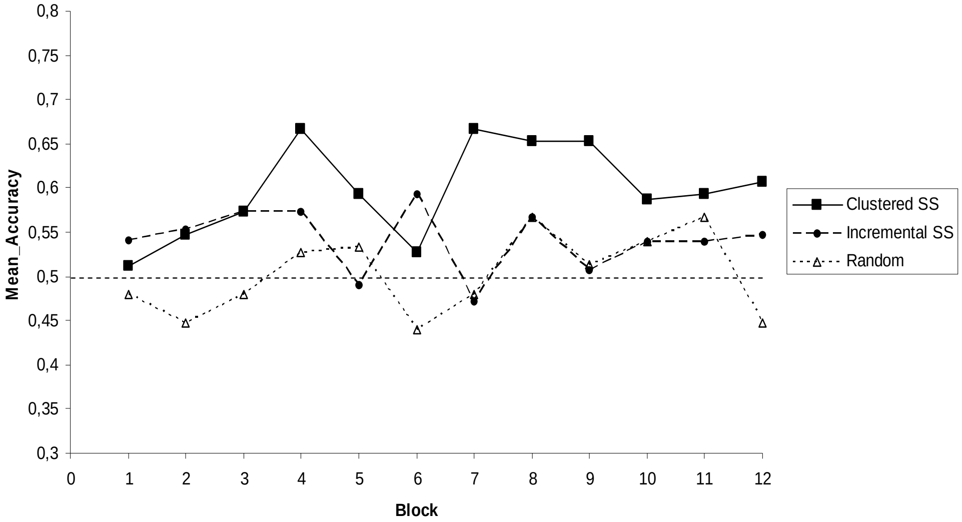


Figure 2. Performance in three groups. The dotted line represents chance level (M = .50). Error bars indicate standard error of the mean.
Figure 2.a is for Experiment 1, and Figure 2.b is for Experiment 2.

To further exclude the possibility that participants might have only concentrated on the outer AB pairs, we compared the performance on two types of violation in 2-LoE test items, (i.e. violations at the last position and violations at the second-to-the-last position) for the clustered SS group. A paired t-test showed that there was no significant difference between the violations at the last position (M= .55, SE= .06), and the violations at the second-to-the-last position (M = .51, SE= .05), *t* (14) = .84, n.s., indicating no particular focus on the first-last positions.

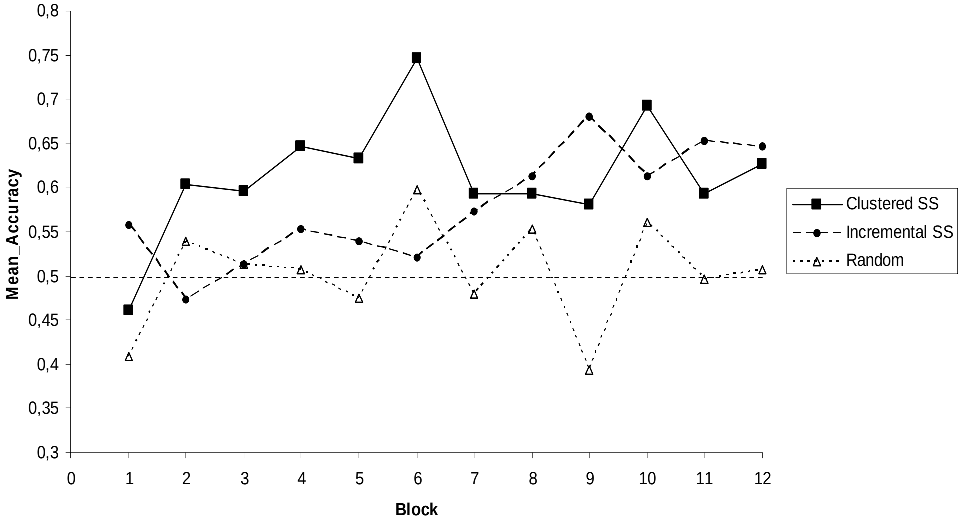
We conducted an additional analysis over performance by block. There was no main effect of block, *F* (11, 462) = 1.29, n.s., nor significant interaction between block and condition, *F* (22, 462) = 1.20, n.s. As shown in Figure 3a, the clustered SS group showed a gradual learning curve.

a)



0,8
0,75
0,7
0,65
0,6
0,55
0,5
0,45
0,4
0,35
0,3
Mean_Accuracy
0
1
2
3
4
5
6
7
8
9
10
11
12
Block
Clustered SS
Incremental SS
Random

b)



0,8
0,75
0,7
0,65
0,6
0,55
0,5
0,45
0,4
0,35
0,3
Mean_Accuracy
0
1
2
3
4
5
6
7
8
9
10
11
12
Block
Clustered SS
Incremental SS
Random

Figure 3. Performance on 12 blocks for three groups. The dotted line represents chance level (M = .50).

Figure 3.a is for Experiment 1, and Figure 3.b is for Experiment 2.

The higher performance in the clustered SS regimen replicated the SS effect in Lai and Poletiek (2011). However, the data regarding the incremental SS group suggested that participants were not assisted by the SS input when it increased gradually rather than discretely in complexity. One possible explanation is that as a consequence of the incrementally growing SS presentation, participants lacked sufficient preliminary training with 0-LoE exemplars only. Indeed, under the incremental SS condition, 1-LoE exemplars were introduced in the second block already, which was before all possible unique 0-LoE items could have been learned.

Possibly, the poor performance in the incremental SS condition was not caused by the incremental format per se, but may have been due to the learners having been deprived of preliminary elaborate exposure to a cluster of the 0-LoE exemplars of the grammar only. Learners started processing recursive loops before they could have acquired solid knowledge of the basic 0-LoE pairs. Their knowledge of the basic pairs might not have been sufficient to detect grammatical 0-LoE pairs in longer items with multiple pairs.

In Experiment 2, we therefore re-conducted Experiment 1 with a skewed frequency distribution of the input items. The frequency distribution was determined according to the probabilities of the unique sequences in a random output generated by the grammar. This output typically produces short items with high probability; long and complex items with low probability (see also Poletiek & Wolters, 2009). Item probabilities were calculated by “running” a statistical version of the grammar (Charniak, 1993). In accordance with this distribution, more 0- than 1- and 2-LoE items would be presented during training (Appendix B).

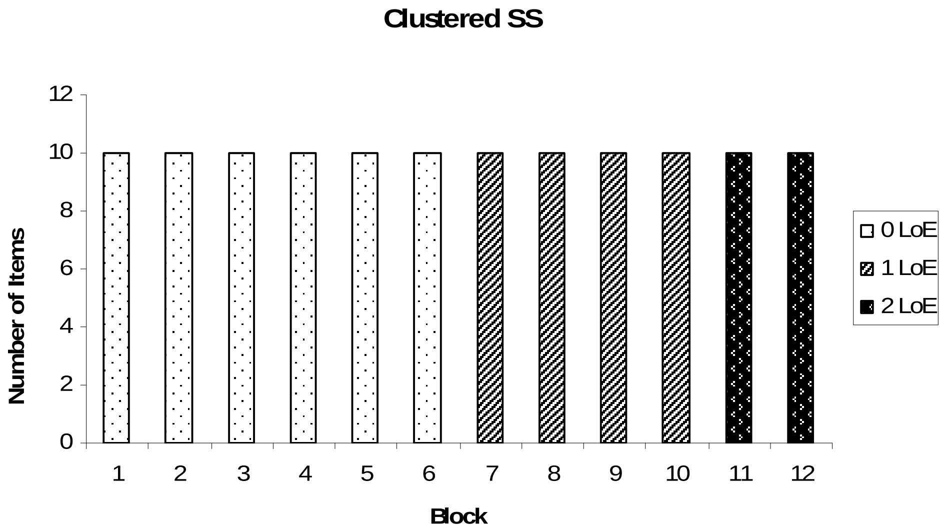
Experiment 2

Method

Participants. Forty-five students from Leiden University participated. None had participated in Experiment 1.

Materials and design. Three experimental groups were presented with 60 items with 0-LoE, 40 items with 1-LoE and 20 items with 2-LoE (Figure 4): The clustered SS group would see 0-LoE only in the first six blocks, 1-LoE items in the next four blocks, and 2-LoE items in the last two blocks. For the incremental SS group, in the first three blocks participants would see 0-LoE items only; In Block 4, two items with 1-LoE were introduced, and gradually, the input would contain more items with higher LoE. For the random group, the same materials were presented randomly.

Importantly, the same test items as in Experiment 1 were used.



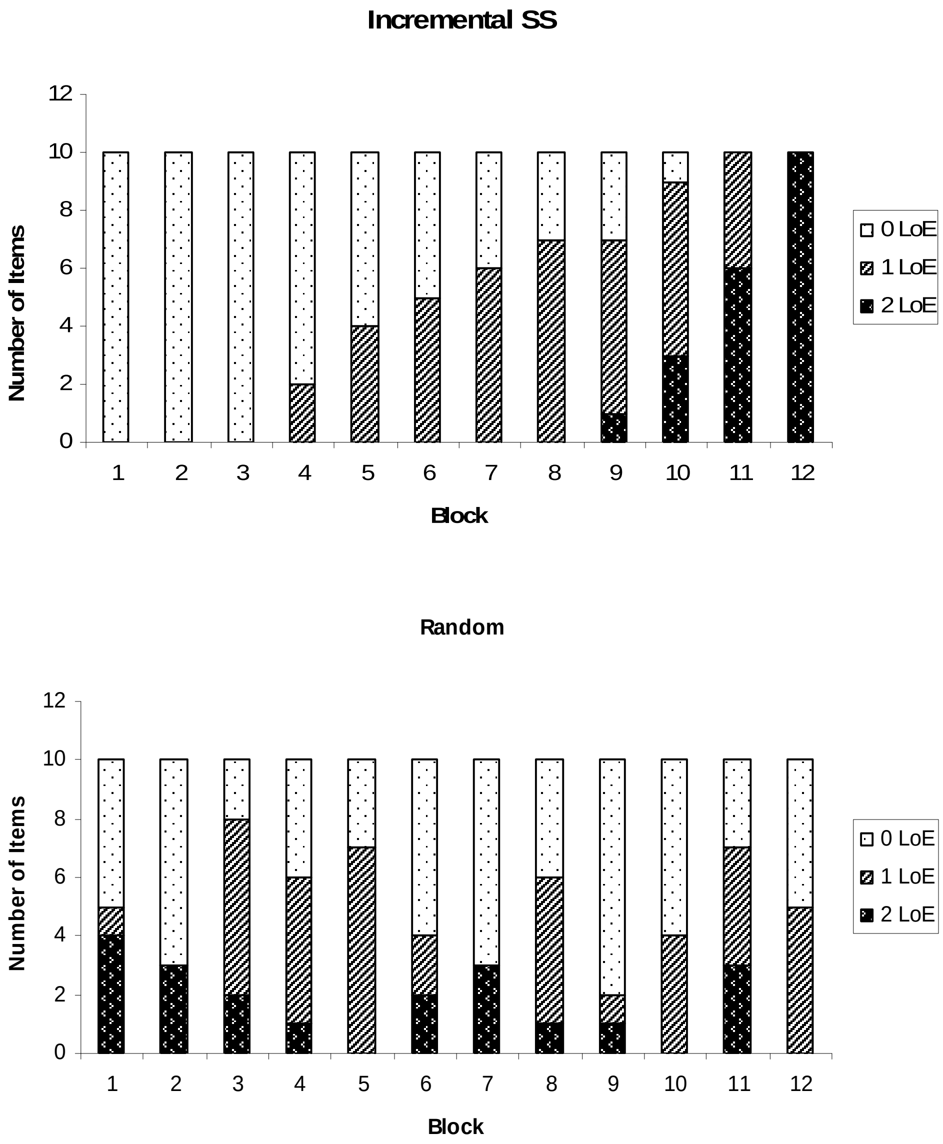


Figure 4. Experiment 2. The ordering of exemplars with 0-, 1-, and 2-LoE in the input under the clustered SS, the incremental SS, and the random condition.

Procedure. Identical to Experiment 1.

Results and discussion

An ANOVA showed a main effect of condition, *F* (2, 42) = 3.90, $p < .05$, $\eta_p^2 = .16$. As displayed in Figure 2.b, performance was significantly better than chance for both the clustered SS group (M = .61, SE = .04), *t* (14) = 3.11, $p < .01$, $r = .64$, and the incremental SS group (M = .58, SE = .03), *t* (14) = 2.39, $p < .05$, $r = .54$. The random group (M = .50, SE = .01) did not differ significantly from chance, *t* (14) = .47, n.s. T-tests indicated significant differences between the clustered SS group and the random group, *t* (28) = 2.97, $p <.01$, $r = .49$, and also between the incremental SS group and the random group, *t* (28) = 2.25, $p < .05$, $r = .39$, but not between the clustered SS group and the incremental SS group, *t* (28) = .73, n.s.

The higher than chance accuracy of grammaticality-judgment in the clustered SS group once again verified the original SS effect in Lai and Poletiek's (2011) study. In addition, in contrast to the results in Experiment 1, the incremental SS group, with a preliminary exposure to three blocks with 0-LoE only, now outscores chance level.

We also compared performance on different types of 2-LoE ungrammatical items. We found no difference between the violations at the last position and at the second-to-the-last position, for the clustered SS group. $M_{\text{Last}} = .64$, $SE_{\text{Last}} = .05$, $M_{\text{Second-to-the-last}} = .62$, $SE_{\text{Second-to-the-last}} = .05$, *t* (14) = .32, n.s., as well as for the incremental SS group, $M_{\text{Last}} = .53$, $SE_{\text{Last}} = .05$, $M_{\text{Second-to-the-last}} = .51$, $SE_{\text{Second-to-the-last}} = .05$, *t* (14) = .48, n.s.

A repeated-measure analysis showed that there was a main effect of block, *F* (11, 462) = 2.80, p < .005, $\eta_p^2 = .06$, and a significant interaction between block and condition, *F* (22, 462) = 2.10, $p < .005$, $\eta_p^2 = .09$ (Figure 3b).

Combined analysis

We probed into accuracy after exposure to various numbers of blocks with only 0-LoE learning items in both experiments (Figure 5). In line with our proposal, mean performance shows an increasing trend correlating with the number of training items with only 0-LoE items presented at the beginning of exposure to the input. When participants were trained with only one block of 0-LoE learning items in the beginning (i.e. the incremental SS group with equal distribution), their performance did not differ from chance, M = .54, SE = .03, *t* (14) = 1.60, n.s. However, when they were exposed to three (i.e. the incremental SS group with unequal distribution), four (i.e. the clustered SS group with equal distribution) or six blocks (i.e. the clustered SS group with unequal distribution) with only 0-LoE learning items, they performed significantly above chance level, M = .58, SE = .03, *t* (14) = 2.39, *p* < .05, *r* = .54; M = .60, SE = .04, *t* (14) = 2.64, *p* < .05, *r* = .58; M = .61, SE = .04, *t* (14) = 3.11, *p* < .01, *r* = .64, respectively.

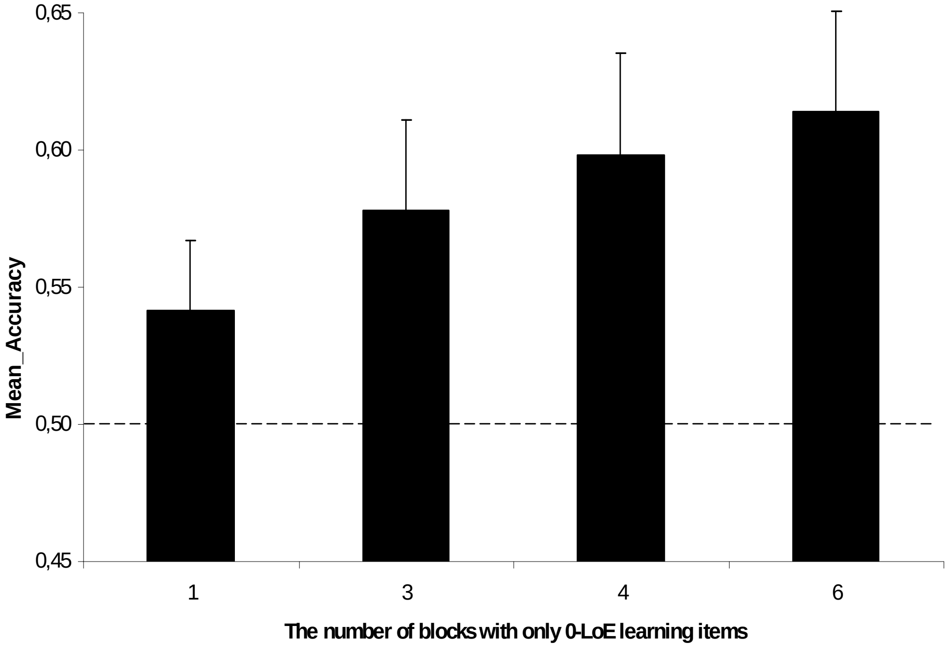


Figure 5. Performance after exposure to various numbers of blocks with only 0-LoE learning items in Experiment 1 and 2. The dotted line represents chance level (M = .50). Error bars indicate standard error of the mean.

Interestingly, participants, who received only 0-LoE learning items during the first three blocks, significantly improved performance on 0-LoE (M= .67, SE= .03), 1-LoE (M= .60, SE= .03), 2-LoE (M= .60, SE= .02), respectively, during the subsequent blocks (Block 4-12) than that during the initial blocks (Block 1-3) on 0-LoE (M= .61, SE= .03), *t*

(44) = 2.04, *p* < .05, *r* = .29; 1-LoE (M= .52, SE= .02), *t* (44) = 2.93, *p* < .01, *r* = .40; 2-LoE (M= .52, SE= .02), *t* (44) = 2.28, *p* < .05, *r* = .33. By contrast, participants, who received all LoE learning items from the beginning, did not improve during Block 4-12 (M= .52, SE= .01), compared to their performance during Block 1-3 (M= .48, SE= .02), *t* (29) = 1.48, n.s. The results indicate that three consecutive blocks with only simple 0-LoE learning items are indeed crucial to grasp the embedding structure displayed in the more complex stimuli that follow.

General discussion

In Experiment 1, we compared the effect of two types of SS training regimens on learning a center-embedded grammar: a discrete ordering with consecutive clusters with increasing LoE for each cluster, and a continuous ordering, in which exemplars with more embedded clauses are gradually inserted in the training input. Only the discrete SS group outperformed the randomly ordered control group significantly. This result replicates the facilitation effect of a discrete SS training regimen in Lai and Poletiek (2011).

The absence of the beneficial effect in the continuous version of the SS training was explained by the absence of sufficient preliminary training on exemplars of the grammar without applications of the center-embedded rule. In Experiment 2, we tested this possibility with the same ordering conditions as Experiment 1, but with exemplars' frequencies inversely related to their complexity (50% of the learning set is simple 0-LoE items, 33% 1-LoE items, and 17% 2-LoE items). The skewed frequency distribution formally corresponds to a random output of the grammar, and resembles the distribution of natural language input, in which short and simple sentences occur more often than long and complex ones. Testing two SS regimens with the skewed distribution provided the possibility to disentangle the contributions of the input ordering from the influence of early exposure to simple items only, on learning. Indeed, when the input distribution was skewed, the learner in the continuous SS condition would still be exposed to a substantial cluster with the simplest exemplars only in the beginning of training.

As we observed, learning was also enhanced by a continuous SS training, when the distribution of the input was skewed to favor highly frequent 0-LoE items. The contribution of the skewed frequency distribution might originate mostly from the massive exposure to 0-LoE items, instead of the decrease of multiple LoE items. Moreover, our proposal that this combined facilitation was accounted for by early intensive exposure to a cluster of 0-LoE only was supported by a finding emerging from both experiments: Only participants, who were exposed to at least three blocks of simple structures without embeddings, showed any learning of subsequently presented center embedded structures. Those presented with embedded items right away did not improve after the first three blocks.

In sum, our data replicate the SS effect also when the input grows continuously rather than discretely from simple towards more complex, but only when the frequency distribution of the exemplars at training favors high numbers of simple exemplars. In this manner, by adapting two characteristics of the input to make it more representative for the natural linguistic input – continuous SS and skewed frequency distribution – we could show how these characteristics of the environment form an optimal setting for learning to emerge.

It seems that the earliest stage of training serves as an essential stepping stone for eventual acquisition of the complex center-embedded grammar. A possible cognitive explanation of this facilitation process is that frequent and early exposure to the basic pattern of the grammar splits up the learning in different consecutive parts with separate learning goals: first the solid acquisition of a basic pattern, and second detecting the recursive operation that operates on that basic pattern. An environment that separates the steps and organizes their time course accurately fits the needs of the learner. As in natural language, the child-directed speech contains mostly shorter and simpler phrases than adult speech (Pine, 1994).

Another aspect of this fit between environment and learner might be the constrained cognitive capacity of the learner in the first stage of exposure. As Kersten and Earles (2001) indicate, the limitation of children’s processing ability made them focus on small constituents first and enabled the ultimate better language performance than adults.

Young children learning language naturally start to process linguistic stimuli using the simplest “model” that accounts for the input (Chater & Vitányi, 2002). The *less is more* (Newport, 1990) hypothesis reflects this idea. First, linguistic sequences may be processed by an associative linear learning mechanism. As the input grows in complexity, along with cognitive capacities, processing might become more complex and hierarchical. Since our study was based on adult participants, future research is needed to investigate how developmental cognitive factors interact with the environmental characteristics investigated here.

Although the current results reveal some crucial properties of the learning process, there are of course limitations of this type of artificial grammar learning studies. For instance, our work mimicked some ideal “error free” learning environment, and used visual materials (Conway et al., 2003; Lai & Poletiek, 2011). Also we tried to simulate the development of children’s learning by observing adults’ behavior in a laboratory task. And we used a fixed artificial meaningless vocabulary, which differs largely much from the rich natural language vocabulary (Fedor, Varga, & Szathmary, 2012). However, there are also undeniable strengths of the artificial grammar learning approach, such as the possibility to investigate the hypothesized factors in isolation, disregarding temporarily the richness of natural language, e.g. semantics (Gomez & Gerken, 1999).

Our experiments indicate that in the lab and possibly in natural learning situations, learners can utilize complexity-based ordering and frequency variations of stimuli over time, as cues to abstract complex pattern information, avoiding in this manner the difficulty of inducing these complex structures by computation.

Footnote

[1] In order to avoid easy detection with the hint of surface heuristics, no repetition of exactly the same syllable was allowed in the same string. In the test string, the number of A's and B's is equal.

[2] This criterion results in: 1) for 1-LoE, the violations would always appear at the last position (e.g. $A_1A_2B_2\underline{\mathbf{B_3}}$); 2) for 2-LoE, we equally divided ungrammatical items into two types of violations: one type with violations at the last position (e.g. $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$), and the other with violations at the second-to-the-last position (e.g. $A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$).

Appendix A

Learning stimuli in the clustered starting small (SS) and the incremental SS condition of Experiment 1

Blocks	Condition	
	Clustered SS	Incremental SS
Block1	detu	detu
	deto	deto
	ditu	ditu
	dito	dito
	bepu	bepu
	bipu	bepo
	bipo	geku
	geku	geko
	geko	giku
	giku	giko
Block 2	detu	detu
	deto	ditu
	ditu	bepu
	bepu	bepo
	bepo	bipo
	bipu	geko
	bipo	giku
	geko	giko
	giku	degekutu
	giko	didetotu
Block 3	detu	detu
	deto	deto
	dito	ditu
	bepu	dito
	bepo	bepu
	bipu	bipu
	geku	giku
	geko	digekotu
	giku	bibepopu
	giko	gebipoku
	deto	deto
	ditu	bipo
	dito	geku
	bepu	geko

Block 4	bepo	giku
	bipo	giko
	geku	dedituto
	geko	bebipupo
	giku	gebepuku
	giko	giditoko
Block 5	dedituto	deto
	degekutu	bepu
	didetotu	bepo
	digekotu	bipu
	bebipupo	geku
	bidetupu	degekoto
	bibepopu	digikotu
	gebepuku	bebipopu
	gebipoku	bidetupu
	giditoko	gedituku
Block 6	debepotu	dito
	degekoto	bipu
	dibeputo	geko
	digikotu	debepotu
	bebipopu	dibeputo
	biditupu	biditupu
	bigekupu	bigekupu
	gedituku	gegikuko
	gegikuko	gidetuko
	gidetuko	didegekutotu
Block 7	debiputu	bipo
	degikutu	degikutu
	digekuto	digekuto
	bedetopo	bedetopo
	begikopu	biditupo
	biditupo	gedetuko
	gedetuko	gibepuko
	gebepoko	debegekoputo
	gibepuko	bidibepotopu
	gigekoku	gibigekopoku
Block8	debipoto	debiputu
	didetuto	begikopu
	dibiputo	gebepoko
	beditopu	gigekoku
	begikupo	biditopo

	biditopo	dedibepututo
	bigikopo	digibepokotu
	gidetoku	begedetukupo
	gibipuku	bidegikotupo
	gigekuko	gidebepotoku
Block 9	dedibepututo	dibiputo
	degebipukotu	beditopu
	dibigikuputo	gidetoku
	digibepokotu	bigikopo
	bedigikutopo	degebipukotu
	bidegikotupo	dibigikuputo
	bigeditokupo	bedigikutopo
	gebedetopuko	bigeditokupo
	gibegekupoko	gebedetopuko
	gibigekopoku	gibegekupoko
Block 10	debegekoputo	didetuto
	degibipokuto	begikupo
	dibigikoputu	gigekuko
	bededitotupu	degibipokuto
	begedetukupo	dibigikoputu
	bidibepotopu	bededitotupu
	bigiditukopu	bigiditukopu
	gebeditopoko	gebeditopoko
	gidibiputuku	gidibiputuku
	gigebipokoku	gigebipokoku
Block 11	debegikopotu	debipoto
	didegekutotu	gibipuku
	digedetokutu	debegikopotu
	bedegikotopu	digedetokutu
	begeditukopu	bedegikotopu
	bibegekupupo	begeditukopu
	gedebeputuku	bibegekupupo
	gebidetupoko	gedebeputuku
	gidebepotoku	gebidetupoko
	gibiditopuko	gibiditopuko
Block 12	debigikupotu	debigikupotu
	dibegekopoto	dibegekopoto
	digidetukuto	digidetukuto
	bedibipotupu	bedibipotupu
	begibipukopo	begibipukopo
	bigedetokupu	bigedetokupu

	gedibipotoku	gedibipotoku
	gebibepupoku	gebibepupoku
	gidebepotoko	gidebepotoko
	gigeditukoku	gigeditukoku

Appendix B

Learning stimuli in the clustered starting small (SS) and the incremental SS condition of Experiment 2

Blocks	Condition	
	Clustered SS	Incremental SS
Block1	detu	detu
	deto	ditu
	ditu	dito
	dito	bepu
	bepu	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	geko	giku
	giku	giko
Block 2	detu	deto
	deto	ditu
	ditu	bepu
	dito	bepo
	bepo	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	geko	giku
	giko	giko
Block 3	detu	detu
	deto	deto
	ditu	dito
	bepu	bepu
	bepo	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	giku	giku
	giko	giko
	detu	detu
	deto	ditu
	dito	dito

Block 4	bepu	bepo
	bepo	bipu
	bipu	bipo
	bipo	geku
	geko	giko
	giku	gigekoku
	giko	gebepuku
Block 5	detu	detu
	ditu	deto
	dito	bepu
	bepu	bipu
	bepo	geko
	bipu	giku
	geku	dedituto
	geko	digikotu
	giku	biditupu
	giko	gigekuko
Block 6	deto	ditu
	ditu	dito
	dito	bepo
	bepu	bipo
	bepo	geku
	bipo	degikutu
	geku	didetotu
	geko	bidetupu
	giku	gibipuku
	giko	begikopu
Block 7	dedituto	deto
	degekutu	ditu
	didetotu	bepu
	digekotu	giko
	bebipupo	degekutu
	bidetupu	digekotu
	bibepopu	bebipupo
	gebepuku	bibepopu
	gebipoku	gebipoku
	giditoko	giditoko
	debepotu	deto
	degekoto	dito
	dibeputo	giku
	digikotu	debepotu

Block8	bebipopu	degekoto
	biditupu	dibeputo
	bigekupu	bebipopu
	gedituku	bigekupu
	gegikuko	gedituku
	gidetuko	gidetuko
Block 9	debiputu	detu
	degikutu	bepo
	digekuto	geko
	bedetopo	debiputu
	begikopu	digekuto
	biditupo	bedetopo
	gedetuko	biditupo
	gebepoko	gegikuko
	gibepuko	gibepuko
	gigekoku	gebeditopoko
Block 10	debipoto	bepo
	didetuto	gedetuko
	dibiputo	dibiputo
	beditopu	gidetoku
	begikupo	bigikopo
	biditopo	beditopu
	bigikopo	debipoto
	gidetoku	didegekutotu
	gibipuku	dedibepututo
	gigekuko	bedegikotopu
Block 11	dedibepututo	didetuto
	degibipokuto	begikupo
	didegekutotu	biditopo
	digibepokotu	gebepoko
	bedegikotopu	degibipokuto
	bidegikotupo	digibepokotu
	bigeditokupo	bidegikotupo
	gedebeputuku	bigeditokupo
	gebeditopoko	gedebeputuku
	gidebepotoko	gidebepotoko
Block 12	debigikupotu	debigikupotu
	dibegekopoto	dibegekopoto
	digedetokutu	digedetokutu
	bededitotupu	bededitotupu
	begedetukupo	begedetukupo

	bigiditukopu	bigiditukopu
	gebedetopuko	gebedetopuko
	gigebipokoku	gigebipokoku
	dibigikuputo	dibigikuputo
	gidibiputuku	gidibiputuku

Appendix C

Testing stimuli in all conditions of Experiment 1 and 2.

Blocks	Grammaticality	
	Grammatical	Ungrammatical
Block1	detu	be**to**
	dibiputu	gedeto**tu**
	bigekupo	gidito**pu**
	degebipokotu	dibigiko**tu**to
	gibidetopuku	bibeditupu**to**
Block 2	ditu	bi**tu**
	debeputo	geditu**po**
	bedetopu	gibepu**to**
	didebepututo	begedetu**to**pu
	gedegikotoku	bigidituko**ku**
Block 3	bepu	ge**tu**
	digikutu	dedito**pu**
	bigekopo	gibepo**tu**
	degibipokuto	digibipo**to**tu
	bigedetukupu	gedibepoto**pu**
Block 4	bipu	gi**to**
	degekotu	bedetu**ku**
	dibipotu	gegiko**po**
	didegikototu	dedibepoto**ku**
	bidegikotopu	gegibipu**to**ku
Block 5	geku	de**pu**
	dibipoto	bidetu**ko**
	beditopo	gibipo**pu**
	dibegekupoto	debegeku**to**tu
	gebedetupoko	gibedetupo**to**
Block 6	giku	di**po**
	degikotu	begiko**to**
	gedetoko	bideto**tu**
	dibidetoputu	debegekopo**ku**
	bedigekotopo	bidegiku**po**pu
	deto	be**ku**

Block 7	gebipuku	digiko**pu**
	gibepoku	bigiku**tu**
	dibibepupotu	debegeko**ku**tu
	gibeditupuko	gebidetopu**tu**
Block8	dito	bi**ko**
	gebepuko	degeku**po**
	gidituku	begeko**tu**
	digebepokutu	degebepoku**ko**
	begiditukopu	bidigeko**po**pu
Block 9	bepo	ge**po**
	bigikupu	debipo**ko**
	gidetoko	digeku**po**
	digebipokoto	degibipu**tu**to
	bigeditokupu	gibibepupo**to**
Block 10	bipo	gi**pu**
	begikupu	debipu**ko**
	geditoko	dibepo**ku**
	bidegikutopo	degibipuko**ku**
	gibegekupoko	gedigiku**po**ko
Block 11	geko	de**ko**
	beditupo	dibepu**ku**
	gidetuku	bidito**ku**
	bibegekupopu	dibedetupo**ko**
	gidebeputuku	gebeditu**to**ko
Block 12	giko	di**ku**
	debepoto	begeku**ko**
	bigekopu	gebipu**to**
	begibipukopo	bidigiku**ko**pu
	gegiditukoku	gigeditoku**po**

Chapter 4

Why we do understand *the dog that the man walks barks* but struggle with *the dog walks the man that barks*:

A Semantic Memory Account for Hierarchical and Linear Linguistic Recursion (SMR)

This chapter is based on: Lai, J. & Poletiek, F. (submitted).

Abstract

Previous theoretical “locality” accounts (Gibson & Thomas, 1996; Kimball, 1973; Lewis, 1996) explain the difficulty of processing hierarchical center-embedded sentences by working memory limitations hindering accurate linking of the long distance dependencies that center embedded constructions generate. Alternately, sentences with right branching relative clauses with dependencies in nearby positions are easier to process. Although a few studies showed effects of semantic characteristics of related words in complex sentences (Blauberg & Braine, 1974; Christiansen & MacDonald, 2009; Powell & Peters, 1973; Stolz, 1967), it is unclear how positional relatedness interacts with semantic relatedness between words in linear and hierarchical constructions. We present a sentence comprehension study manipulating structure (hierarchical and linear) and the congruency between the semantic and positional pattern of word associations (match, mismatch and neutral) in the sentence. The data suggest a strong influence of semantic-syntactic pattern congruency, which occasionally even fully overshadowed difficulties caused by syntactical structure and positional distance. Moreover, this congruency effect was equally strong for linear and for hierarchical structures. We propose our semantic-memory model for processing recursive (SMR) structures to account for this effect, which can not be explained by the classical locality view. SMR also challenges the classical assumption that hierarchical structures are complex and linear not (Gibson, 1998).

Sentence complexity has been a notable focus of interest to psycholinguists. Recently, linguistic recursive complexity has been proposed to be the crucial factor distinguishing humans and nonhumans (Bloomfield, Gentner, & Margoliash, 2011; Corballis, 2007; Grainger, Dufau, Montant, Ziegler, & Fagot, 2012; Hauser, Chomsky, & Fitch, 2002; Lai & Poletiek, 2011; Rey, Perruchet, & Fagot, 2012). Recursion is a computational self referential mechanism, which allows for a finite number of rules to produce an infinite set of output (Chomsky, 1957). There are many types of recursive rules in language. However, one particularly complex type of recursion in natural language sentences has been much studied, namely, center-embedded (CE) structures, typically described formally as A^nB^n grammar (Fitch & Hauser, 2004). Assuming two word categories: A-words (e.g. nouns in natural language) and B-words (e.g. verbs in natural language), the CE A^nB^n grammar specifies a basic rule about which A words may be paired with which B words, and a recursive operation for inserting a grammatical AjBj pair within another AiBi pair to result in a new grammatical sentence. In this manner, CE sentences follow an AiAj…BjBi pattern. Since the embedding structure involves a "stack" of syntactically dependent elements possibly far away from each other in the sentence (e.g., Ai and Bi), CE structures are called "hierarchical" and non-linear and therefore require hierarchical cognitive processing (Christiansen & Chater, 1999). For example, in the natural sentence (1) with a CE structure, a higher order non-linear process of binding each A's to a specific B is required for correct comprehension.

Recursive rules can be linear, however, as well. In right branching (RB) structures of type $(AB)^n$, in which syntactically related AB pairs are close or even adjacent to each other. In the RB sentence (2), for example, the positional close distance between A and B elements is a direct cue for their syntactical relatedness, facilitating a simple linear parsing strategy.

(1) *John saw that the cat that the dog that the man walked chased ran away*. [CE]
A_1 A_2 A_3 B_3 B_2 B_1

(2) [1]*John saw that the man walked a dog that chased a cat that ran away.* [RB]
A_1 B_1 A_2 B_2 A_3 B_3

The difference between hierarchical structure and linear structure is crucial in linguistic theories on learnability of language (Poletiek & Lai, 2012), and according to recent theorising, parallels the distinction between the human and animal language faculty (Fitch & Hauser, 2004; Hauser et al., 2002). Hierarchical processing has been argued to imply cognitive control, higher order computation, consciousness and executive control, in contrast to linear processing, relying on low level memory and associative mechanisms. However, general cognitive limitations clearly affect and limit the processing of hierarchical structures, as evidenced in our difficulties to parse structures with multiple clauses in natural language, and also in experimental studies on learning complex artificial systems. Though most authors agree that there is some role for working memory mechanisms in processing hierarchical structures, it is still empirically unclear and under debate, *how* memory mechanisms and associative learning come into play, and whether they suffice to account for how human language users deal with hierarchical recursion (Friederici, Bahlmann, Friedrich, & Makuuchi, 2011; Makuuchi, Bahlmann, Anwander, & Friederici, 2009). The present study focuses on these questions. In particular, we argue that memory and associative learning mechanisms can largely explain recursive language processing, if we take into account the semantic aspects of the linguistic input and the way our memory deals with semantically rich content. By assuming memory content to be meaningful, we can push the working memory account to a clearer and more powerful explanation of complex linguistic behaviour.

Previous research has concentrated on the problem of limited memory as a quantity, i.e. memory load and computational integration effort have been argued to affect

---

[1] In the Dutch translation of the RB sentence (2) slight positional changes would occur, because the object can move in front of the verb in the relative clause (being "verb final"). However, the typical contrast between short distances for the linear RB constructions and the long distances in the CE constructions are conserved in Dutch. Also, the RB clauses are lined up in a linear sequence, over time, in both languages. See also Appendix. The Dutch translation of sentence (2), and the word-by-word translation back in English are:

(2) Jan zag dat [de man een hond uitliet] [die een kat achtervolgde] [die wegrende]
John saw that [the man a dog walked] [that a cat chased] [that ran away].
A_1 A_2 B_1 A_3 B_2 B_3

differentially CE and RB sentence comprehension, because items have to be kept in memory simultaneously and for a longer period of time in the former than in the latter structure. For example, CE sentences normally require retaining a certain subject noun in working memory until it can be associated with its further located predicate (verb), whilst in the meantime additional noun-verb pairings have to be determined. In sentence (1), “the cat” has to be encoded, stored and retrieved from memory, when “ran” appears at the end of the sentence. In the middle of the sentence, two other nouns have to be stored and retrieved, though not before the associated verb shows up.

It is not surprising then, that a large number of studies suggest that CE sentences are more difficult to understand than their RB counterparts (Bach, Brown, & Marslen-Wilson, 1986; Blauberg & Braine, 1974; Blumenthal & Boakes, 1967; Caplan & Hildebrandt, 1988; Christiansen & Chater, 1999; Fodor & Garrett, 1967; Gibson & Thomas, 1999; Hildebrandt, Caplan, & Evans, 1987; Larkin & Burns, 1977; Marks, 1968; Miller, 1962; Miller & Isard, 1964; Poletiek, 2011). Moreover, it is generally assumed that this difficulty increases fast with the number of levels of embedding (LoE) for hierarchical structures, since the dependencies are pushed away from each other further with each added clause. For RB sequences, LoE is thought to weakly affect difficulty or not at all (Chomsky, 1965; Church, 1982; Gibson, 1998; Marcus, 1980; Reich, 1969; Stabler, 1994). From 2-LoE on, i.e. two clauses hierarchically nested in the main clause, sentences are barely understandable (de Vries, Petersson, Geukes, Zwitserlood, & Christiansen, 2012; Foss & Cairns, 1970; Miller, 1962; Vosse & Kempen, 1991). This increasing complexity is reflected in its occurrence in actual natural languages: 2-LoE sentences are rare in written and even rarer in spoken language (Karlsson, 2010). The complexity level that an actual language user may have to deal with in natural CE sentences, therefore, ranges from 1- to 2-LoE.

Theories explaining the difficulty to process CE sentences and the relatively low accuracy in comprehending them have pointed at memory capacity constrains and computational limitations. The *structural configuration account* (Chomsky, 1965; Miller & Isard, 1964) suggests that the low acceptability of CE sentences is due to the manner in which these self-embeddings are configured. Since human parsers must apply the mirror-like recursive operation to process each clause, they have to remember "re-entries" of each previous clause to reach the highest level (Holmes, 1973). This unique configuration of dependencies can increase in complexity beyond the human computational capacity (Johnson, 1998). Just and Carpenter's (1992) *working memory theory of comprehension* is based on a similar reasoning that complex structures require more integration and memory resources, explaining differences in processing difficulty for CE and RB constructions, but also individual differences.

A more recent account, *the processing overload account* (Gibson & Thomas, 1996; Kimball, 1973; Lewis, 1996) proposes that the integration of corresponding elements into one constituent costs more cognitive resources in CE than in RB structures, which allow for immediate integration of syntactically related elements thanks to the adjacent locations mirroring their syntactical relatedness. The *syntactic prediction locality theory* (SPLT) by Gibson (1998) provides further theoretical refining of this working memory account. The SPLT proposes that locality has a strong impact on both the integration cost and memory cost: For integrating, the computational resources needed to connect two related events increase along with the number of constituents to be related in the sentence and the distance between them. Regarding memory costs, it requires more capacity to maintain a local word in memory during a longer period of time before it can be associated with its counterpart. Summing up the common features of theories explaining differential processing difficulties for linear and hierarchical recursive constructions, it is assumed that computational and memory load increase for parsing CE sentences as compared to RB ones, because of the complex association pattern of the elements and the long distance between them in a CE sentence. For RB sentences, related elements being close or even adjacent to each other, memory and integration processes are hardly needed.

In line with these theoretical accounts, experimental studies have explored the effect of structure and level of complexity on cognitive processing, using both natural language materials (Bach et al., 1986; Blauberg & Braine, 1974; Kidd, Brandt, Lieven, & Tomasello, 2007) and artificial grammars (Conway, Ellefson, & Christiansen, 2003; de Vries, Monaghan, Knecht, & Zwitserlood, 2008; Fitch & Hauser, 2004; Lai & Poletiek, 2011; Poletiek, 2002; van den Bos & Poletiek, 2010). In Blauberg and Braine's (1974) early study, participants' comprehension of auditory presented CE and RB sentences with increasing LoE were compared. They found that RB sentences were more understandable than CE ones, and higher LoE hindered CE more than RB in comprehension. With 3-5 LoE, RB sentences were hard but still intelligible, while CE sentences became "virtually impossible" beyond 2-LoE. For 1- and 2-LoE sentences, however, accuracy of processing did not differ significantly between RB and CE sentences. Blauberg and Braine concluded that it was the unique hierarchically nested property of CE, which posed obstacles for comprehension.

Findings with the Artificial Grammar Learning (AGL) paradigm are consistent with natural language studies' findings. In the AGL procedure, typically, participants are trained with $A_1A_2A_3…B_3B_2B_1$ sentences (produced by a CE grammar) or $A_1B_1A_2B_2A_3B_3…$ (produced by a RB grammar) depending on the structure tested. After, participants give grammaticality judgments for new strings being either grammatical or ungrammatical. Accuracy of the grammaticality judgements indicates the amount of learning of the underlying grammar. Research using this paradigm suggests that CE structures are more difficult to learn than RB structures (Christiansen & Chater, 1999; Conway et al., 2003). De Vries, Monaghan, Knecht and Zwitserlood (2008) even found no learning at all of the hierarchical nested pattern of CE structures in an artificial grammar. However, recent studies have looked at extra linguistic factors that might help; for example, prosodic cues (Mueller, Bahlmann, & Friederici, 2010), frequency of occurrence of different types of CE structures (Reali & Christiansen, 2007), experience with complex grammatical constructions (MacDonald & Christiansen, 2002), animacy of the noun (Mak, Vonk, & Schriefers, 2002, 2006), and a starting small training regimen presenting the exemplars over time in increasing order of complexity, and overtraining with the simplest exemplars (without embeddings) (Lai & Poletiek, 2011; Poletiek & Chater, 2006). These studies revealed that factors external to the positional structure can help the integration and memory processes required to process these sentences.

A poorly attended but very straightforward factor that might support parsing messages with complex dependencies is simply the meaning of these dependencies. The semantic factor has hardly been considered in the discussion about the learnability of hierarchical recursion – considered to be a matter of syntax (Goldberg, 2003). The present work explores the effect of the semantic relations between syntactically dependent words in complex hierarchical sentences, and, for the first time, explores how semantic effects differentially influence non-adjacent hierarchical and linear dependencies. Especially, we model prior knowledge of language users about the semantic relations between words (e.g. A and B words) in terms of semantic distance, in analogy to positional distance. This *semantic* distance might help or hinder comprehending, depending on its congruency with the *syntactic* (i.e. positional) distance between syntactically related elements. In this manner, we explain how semantic features of the syntactically dependent elements affect cognitive processing.

The general influence of semantic effects on syntactical parsing has been shown in a number of studies (Fedor, Varga, & Szathmary, 2012). For example, in the sentence "Mary cut the bread with a knife", the syntactic pairing of "cut" and "knife" point in the direction of the correct syntactic analysis, and comprehension becomes easier (MacDonald, Pearlmutter, & Seidenberg, 1994). In an early study, Slobin (1966) showed a similar effect for parsing passive voice sentences: when the relations between two nouns were indeterminate, i.e., object and subject were reversible according to real world knowledge (as in *the girl is being held by the boy*), comprehension was more difficult than when they are irreversible (*the baby is being held by the mother*). In the same vein, Gennari and MacDonald (2008) compared processing of English sentences with objective relative- and subject relative-clauses. They denoted that semantic indeterminacy strongly caused comprehension difficulty.

A few experiments specifically looked at semantic influences in hierarchical CE constructions (Blauberg & Braine, 1974; Powell & Peters, 1973; Stolz, 1967), and more recently, computational work with Simple Recurrent Networks has been carried out (Fedor et al., 2012; Rohde & Plaut, 1999). In his early study, Stolz (1967) exposed his participants to 2-LoE CE English sentences ($A_1A_2A_3B_3B_2B_1$) and observed comprehension under different conditions of semantic relations between the A's and B's. When the semantics of the syntactically related A's and B's determined their relatedness (e.g., *dog barks*), human decoders "do very little syntactic processing" to find out the *syntactic* correspondences between individual A's and B's (Stolz, 1967). Syntactic analysis only occurs when it is highly necessary for understanding. Powell and Peters (1973) replicated Stolz's findings, and concluded that "semantically supported sentences were easier to comprehend and decode than were semantically neutral sentences" (e.g., *man walks*). Besides these early experimentational studies, a few computational and mathematical models (Poletiek & Lai, 2012; Rohde & Plaut, 1999; Weckerly & Elman, 1992) have looked at the effect of semantic biases. Rhode and Plaut (1999) found better performance for a computational model of CE pattern learning when semantic biases were present in the input. Moreover, Weckerly and Elman (1992) observed different performances for two sets of CE sentences: one set with semantic bias, i.e. verbs which were compatible with specific subjects/objects only, the other set without semantic bias. Training with a semantic biased input led to better performance.

On the basis of results reported thus far, semantically supporting content per se seems to help syntactical parsing of various syntactical constructions, including CE. However, we don't know whether semantics differentially tap into hierarchical structures as compared to "easy" linear structures. Such an interaction would be expected on the basis of a locality view on complex sentence processing. If semantic biases do not differ for RB and CE, then both long distance and short distance constructions might be controlled by semantic memory in same way; and semantic distances rather than the positional distances might determine how we deal with complex grammatical patterns.

Another open question is how *interfering* rather than supporting semantic relations affect CE and RB processing. Past research has only looked at two possible semantic biases: it compared *supporting* semantic cues (determinate) and *neutral* (indeterminate) ones. How do *negative* semantic cues affect recursive sentence processing? This third possibility, in which a syntactical analysis goes against a preferred semantic one, can crucially reveal which role is left over for syntactical analysis when a dominant semantic analysis is available, in linear RB structures versus complex CE hierarchical structures. Less interference is expected for positional easy constructions than for hierarchical constructions, by locality based views. If incongruent semantic content, however, interferes equally with hierarchical and linear constructions, then this speaks against a substantial role for linearity or hierarchy as crucial determinants of cognitive processing, in the presence of a semantic cue, even a cue that goes against the syntactic analysis. We propose a semantic-memory model of recursion (SMR) to deal with these questions. SMR makes specific predictions for processing difficulties across recursive structures, LoE, and semantic features of the clauses. In the same manner as memory for paired words is affected by the semantic relation between the to-be-memorized words, sentence processing of embedded sentences is hypothesized by SMR to be affected primarily by the *semantic* pattern of distances between the words that are to be integrated in the sentence, rather than the positional distances.

Regarding semantic "distances", ever since Craik and Tulving (1975), recall performance of word pairs has been shown to vary highly depending on their semantic "distance". Consider the following pairs of words to be memorized:

1) Dog bites / Girl cries / Bird flies
2) Dog walks / Girl runs / Bird stands
3) Dog cries / Bird bites / Girl flies

The first pattern of word pairs is plausible and determined. The second list is plausible but undetermined, since pairings could equally well be interchanged. The third list is highly implausible (going against the plausible pairings of 1) but determined (according to the alternative pairing pattern 1). We would have no difficulty to produce the second word of each pair in the first list when primed with the first word, but finding the correct pattern of matched words in list 2 and 3 poses much more difficulties. SMR assumes these strong semantic pairing effects on memory performance, rather than "locality" (distance in time and space between the items) to explain recursive sentence processing.

SMR is rooted in a usage based view on processing recursive complex language, assuming general memory and associative cognitive processes to underlie how we deal with linguistic stimuli (Christiansen & Chater, 1999; Perruchet & Rey, 2005; Tomasello, 2000). SMR, however, specifies in detail the process by which non-linguistic general memory mechanisms operate to achieve comprehension of non-linear messages.

Two crucial hypotheses of SMR are tested in our experiment: The first hypothesis is about the role of increasing complexity on processing recursive structures. Though both SMR and locality-based accounts predict that more embeddings lead to more processing difficulty, simply because the list of to be memorized pairs increases, locality theories also predict that this effect is stronger for CE (in which positional distance between dependencies increases along with the number of LoE, and thus multiple items have to be retained for a longer period of time and integrated according to an analysis of their positions) than for RB. In SMR, increased depth of embedding is predicted to affect all types of positional patterns equally, because the memory process resourced to retain and integrate the pairs relies on semantic rather than positional information. Importantly, this prediction of SMR holds for a realistic range of complexity in natural language use, i.e. no more than 2-LoE.

The second hypothesis is about the effect of semantic patterns on positional distances between syntactically related elements. Previous research suggests that when semantic associations are determinate and congruent with syntactic positional associated elements in hierarchical CE constructions, processing is facilitated (Stolz, 1967). For RB linear constructions, where the to-be-paired elements are nearby, no semantic facilitation is needed nor expected by locality-based models, and semantic interference will not substantially affect the analysis, because of the clear positional cue. In sum, if the semantic relations are clearly incongruent with the syntactical associations, the locality view predicts linear RB constructions to be less hindered than hierarchical CE constructions with long distances. SMR, however, predicts semantic facilitation or interference *independently* of positional structure. Our model predicts a strong effect of semantic bias, but no interaction with structure. It is only when semantic relations between syntactic pairs are neutral and fully indeterminate, that elements positioned close to each other might be easier to process than distant ones. We predict on the basis of SMR, this suppressing of syntactical analysis to occur for easy (linear) and “difficult” (hierarchical) constructions equally, when the semantic cues are at odds with the positional pattern.

Experiment

In the present experiment, we manipulate structure (CE and RB), LoE (1- or 2-), and three conditions of congruency between the semantic and the syntactic pattern of words pairing, in sentences with one or two relative clauses. In the "*match*" condition, pairing patterns are congruent; in the "*mismatch*" condition, they are incongruent; and in the "*neutral*" condition, they are indeterminate. The three possibilities are displayed schematically in Figure 1. The following sentences illustrate the CE and RB constructions with all types of semantic-syntactic congruency conditions.

(3) *The dog that the boy pats barks.* [Match-CE]
(4) *The dog that the cat watches runs.* [Neutral-CE]
(5) *The boy that the dog pats barks*. [Mismatch-CE]
(6) *The boy pats the dog that barks.* [Match-RB]
(7) *The cat watches the dog that runs.* [Neutral-RB]
(8) *The dog pats the boy that barks.* [Mismatch-RB]

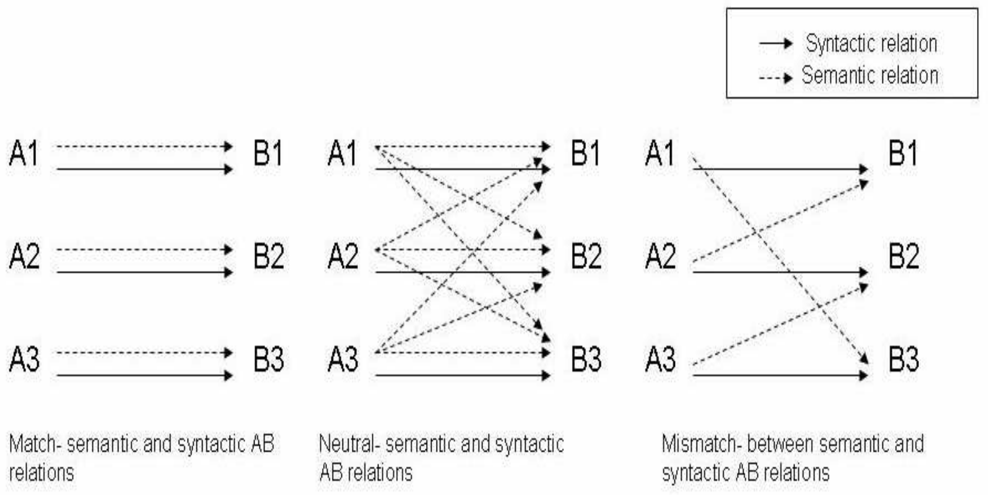


Figure 1. Schematic representation of the Syntactic-Semantic relations between A and B used in the stimulus sentences of type $A_1A_2A_3B_3B_2B_1$.

Method

Participants. Thirty-nine students (34 female), from Leiden University participated in the experiment for course credit or payment. All were native Dutch speakers. All had normal or corrected to normal vision.

Materials and design. There were 96 Dutch stimulus sentences with either one or two relative clauses (see Appendix). For each stimulus sentence, two short test sentences were constructed. Test sentences contained one subject and one predicate only. They served as test of participants' comprehension of the long stimulus sentence displayed previously. One of the short test sentences summarized an event actually described in the corresponding long stimulus sentence. The other one depicted a situation that was *not* a correct description of the content of the stimulus sentence. The incorrect test sentences contained nouns and verbs that were actually present in the stimulus sentence, but in other thematic roles than those in the stimulus sentence. For instance, in an incorrect test sentence, a subject noun could be associated with an unmatched predicate. For example, the stimulus sentence *the girl the dog bites cries* could have *the dog bites* (correct*)*, and *the girl bites* (incorrect) as corresponding test sentences. The short summary sentences could refer to any subject in the long sentences[2]. Two counter-balanced test lists were created to ensure that each stimulus sentence had both a correct and an incorrect short test sentence. Participants were assigned randomly to one of the two counter-balanced lists, which again randomized the ordering of sentences across participants. Proportion of correct responses indicated comprehension accuracy.

The set of stimulus sentences had one of the three possible sentence structures: complex sentences with CE; complex sentences with RB; and simple sentences used as fillers. Since RB relative clauses used in the materials are verb final, subject verb pairs could be not fully adjacent, but separated by an object noun (see Appendix). For example: *Kees zag dat de man(A1) de hond(A2) uitliet(B1) die blafte(B2)* [in word-by-word translation: *Kees saw that the man the dog walked that barked.*] Overall, in our materials, RB constructions could have associated AB pairs separated by one or two words at most (short distance dependencies) and CE constructions could have AB pairs separated by eight words, for 2-LoE sentences. Furthermore, stimulus sentences had one out of three semantic types: *match*, i.e. the syntactical association pattern was congruent with the semantically most plausible association pattern; *mismatch*, i.e. the syntactical association pattern was incongruent with the semantically most plausible association pattern; and *neutral*, the

---

[2] However, for CE stimulus sentences with 2-LoE, A_2B_2 and A_3B_2 test sentences were excluded. This is because in the Dutch sentences used, the subject of B_2 is ambiguous. Grammatically, it can be either A_2 or A_3. See Appendix.

syntactical association pattern was unrelated to any semantic association pattern, because the semantic associations were indeterminate. In summary, the experimental stimulus sentences were manipulated orthogonally according to their structure (RB or CE), according to the match between the syntactical association pattern of A's and B's, and the semantically most plausible association pattern, and according to LoE (see Appendix for an example of each type of stimulus sentence).

Procedure. Participants were seated in front of a monitor, and were instructed that they would be exposed to pairs of Dutch sentences, visually. They would first see a long sentence, and immediately after, a short one. They had to judge whether the test sentence corresponded with the content of the stimulus sentence, or not, by pressing a YES key or a NO key. Participants were required to answer as quickly and as accurately as possible. Each trial started with a fixation cross (500ms) at the center of the screen. Each stimulus sentence began with "Kees weet dat …" (means "Kees knows that …" for 1000ms), and then appears word-by-word (800 ms per word, no interval in-between). It was followed by the short test sentence presented in the same manner. The task took approximately 35 minutes.

Results

In response to recent proposals regarding psycholinguistic data analysis accounting for both variance between participants and item simultaneously (Baayen, 2008; Brysbaert, 2007; Locker, Hoffman, & Bovaird, 2007), the analysis was carried out using a mixed-effects modelling. According to our first main hypothesis, number of LoE (one versus two) affects processing difficulty; and LoE affect processing hierarchical and linear sentences to the same extent. There was a main effect of LoE, F (1, 81) = 23.77, $p < .001$, but no main effect of sentence structure, F (1, 81) = 2.23, n.s., nor a significant interaction between LoE and structure, F (1, 81) = .07, n.s. As displayed in Figure 2, for CE sentences, performance on 1-LoE (M= .84, SE= .02) was significantly better than that on 2-LoE (M= .72, SE= .02), t (38) = 7.02, $p < .001$. Similarly, with RB sentence, performance on 1-LoE (M= .86, SE= .02) was significantly better than that on 2-LoE (M= .76, SE= .02), t (38) = 5.22, $p < .001$. At 1-LoE, performance over CE did not differ from that over RB significantly, *t* (38) = 1.72, n.s.; also, at 2-LoE, the difference between CE and RB did not reach significance, *t* (38) = 2.00, n.s.

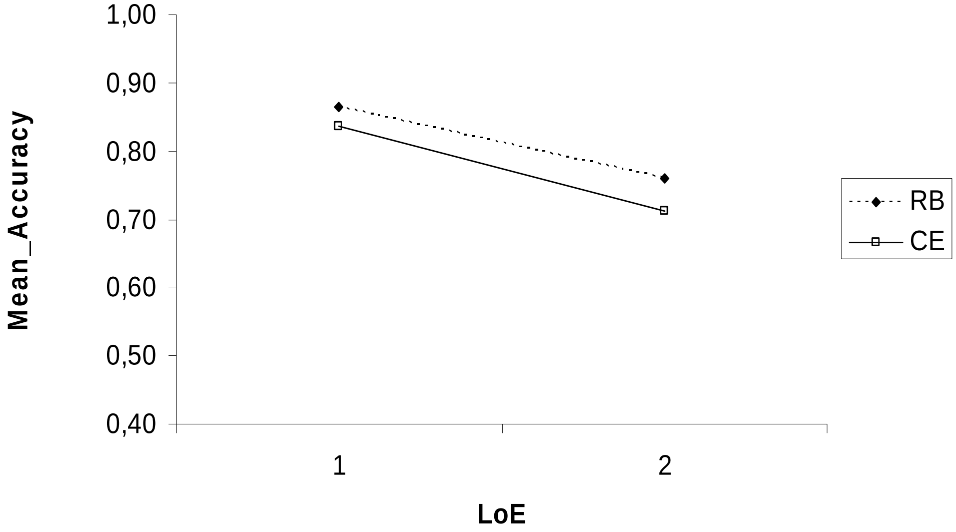


Figure 2. Mean accuracy for RB and CE sentences with 1- and 2-LoE.

Secondly, according to SMR, a strong main effect of semantic-syntactic congruency on accuracy is expected, and no interaction between congruency and structure is expected by SMR, though it is predicted by locality approaches.

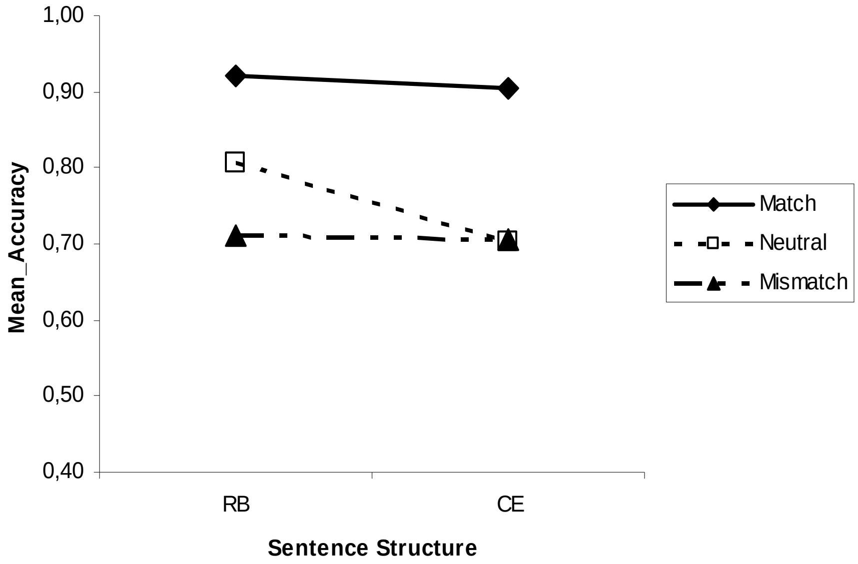


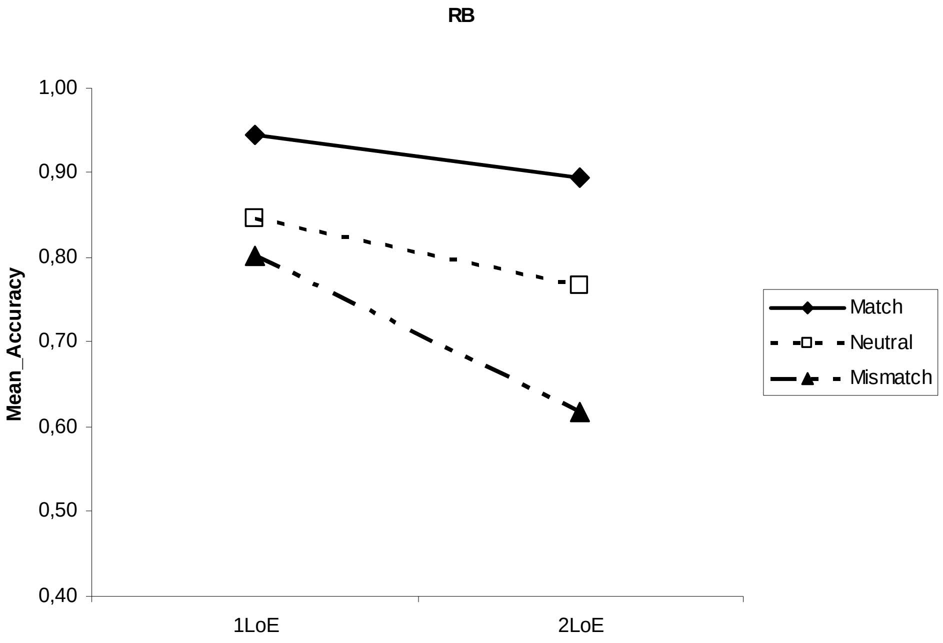
Figure 3. Mean accuracy for RB and CE over three semantic types.

The results indeed show a main effect of semantic type on accuracy, *F* (2, 81) = 31.88, *p* < .001, but no significant interaction between semantic type and structure, *F* (2, 81) = 1.83, n.s. There was no significant three-way interaction (Semantic type × Structure × LoE) either, *F* (2, 81) = 1.16, n.s. (Figure 3). Performance on semantic-matched (congruent) items (M=.92, SE= .01) was significantly better than on semantic-neutral ones (M=.76, SE= .02), *t* (38) = 10.56, *p* < .001, which was better than performance on semantic-mismatched ones (M=.72, SE= .02), *t* (38) = 2.50, *p* < .05.

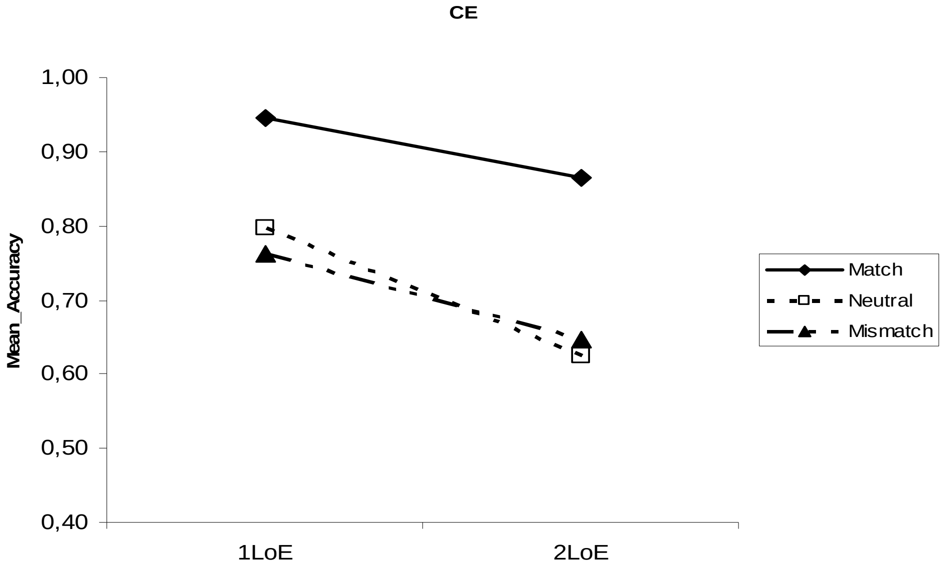
Though the interaction between semantic type and structure was not significant overall, Figure 3 shows differential performance on RB and CE for the neutral items, indicating that only these semantically neutral items were sensitive to positional organization of the pairs. This sensitivity was absent for items with either a matching or mismatching cue. For matched items, RB structures (M= .92, SE= .02) did not differ from CE structures (M= .92, SE= .01), *t* (38) = .04, n.s. Similarly, RB mismatched structures (M= .71, SE= .03) did not differ from CE mismatched structures (M= .72, SE= .02), *t* (38)

= .32, n.s. Only for the neutral items, did performance for RB structures (M= .81, SE= .02) surpass performance for CE structures (M= .71, SE= .02) significantly, *t* (38) = 3.32, *p* < .005 (see also Figure 4).

Figure 4 summarizes the effects of the manipulations taken together; only in the absence of semantic cues, RB constructions outperform CE constructions, and CE constructions are more strongly disrupted by an additional LoE than RB ones. For CE, the difference between mismatch items and neutral ones was not significant, *t* (38) = .33, n.s. Notice further two contrasts displayed in Figure 4 that are inconsistent with a locality view: for CE, 2-LoE matched sentences (M= .88, SE= .02) were scored even better than CE 1-LoE neutral (M= .79, SE= .03), *t* (38) = 2.82, *p* < .01, or CE 1-LoE mismatched ones (M= .76, SE= .03), *t* (38) = 3.86, *p* < .001. CE sentences with matching semantic-syntactic content, with both 1- and 2-LoE items, were better processed than RB sentences without semantic cue, *t* (38) = 6.42, *p* < .001.



(a)



(b)

Figure 4. (a) Mean accuracy for 1-, and 2-LoE RB over three semantic types.
(b) Mean accuracy for 1-, and 2-LoE CE over three semantic types.

Discussion

The present sentence comprehension study compares, for the first time, the effects of positional and semantic aspects of dependencies, in linear RB versus hierarchical CE structures, putting our SMR against the standard locality view. Our results show that for recursive sentences within the range of complexity that is actually present in natural language (1- or 2-LoE), sentence structure did not affect comprehension. Two levels of embedding sentences were more difficult to process than one level sentences, however. But this effect was independent of structure. Thus, the detrimental effect of additional relative clauses (that directly affects sentence length) was not larger for hierarchical structures where the dependencies are pushed apart to further positions, than for linear structures where the dependent elements remain in constant nearby positions. This data is hard to explain by a locality perspective that predicts more difficulties for multiple embeddings in hierarchically organized recursion than in linear recursion.

Furthermore, in line with SMR, there was a strong influence of the preferred semantic association pattern of dependent elements on comprehension. When a semantic cue was available to associate dependencies pair-wise, it strongly facilitated comprehension, if that cue was *congruent* with the *syntactical* association pattern. Inversely, it strongly hindered comprehension if the cue was *incongruent* with the positional association pattern. Strikingly, the semantic cue affected comprehension independently of the sentence structure. When a semantic association scheme for the words was available, it would strongly determine the sentence interpretation, whatever the positional scheme being linear or hierarchical. For example, both the linear sentence *the girl bites the dog that cries* and the hierarchical structured sentence with the same semantic content *the dog the girl bites cries* elicit an inaccurate but semantic plausible interpretation equally often. Another indication of the secondary role of positional information was that number of LoE failed to influence this semantic bias differentially for RB and CE structures. As accounted for by SMR, when there is a clear semantic pairing scheme for the elements, it strongly directs the integration of the sentence, whatever the positional distance of these elements. Positional factors also do not play a greater role with 2- than with 1-LoE.

In line with past findings, when there was no semantic cue to organize and retain in memory the pairing of elements, positional patterns mattered. The “pure” syntactical analysis then performed was remarkably poor, though, varying from 85% accuracy for the easiest linear sentences with 1-LoE to 62% for hierarchical sentences with 2-LoE. Though the latter contrast seemingly supports the locality view, the overall low performance for the neutral sentences when the structure is linear (and thus the positional conditions optimal) remains puzzling, for both locality theories and classical linguistic models assuming a predisposition for parsing the grammars of human languages (Chomsky, 1965; Church, 1982).

Our experimental results challenge the view that comprehension of recursive linear RB structures is generally better than that of recursive hierarchical CE structures (Foss & Cairns, 1970; Marks, 1968; Miller & Isard, 1964), because CE structures require on the one hand the elements to be retained during a longer period of time in memory, and on the other hand, a more sophisticated computational mechanism to determine the paired association of the elements than in RB sentences (Gibson, 1998). Instead, we found that semantic "distance" between the elements actually cause friction to or alternately subserved an accurate analysis. When the semantic association scheme happens to be in line with the syntactic scheme, recursive sentences are processed easily, *whatever* the syntactic scheme. The memory processes that are resourced to achieve comprehension of complex sentences also support our memory for meaningful materials corresponding to real world knowledge, autobiographic and contextual knowledge.

A similar semantic driven mechanism for sentence comprehension was proposed in the "good enough" parsing approach (Ferreira, Bailey, & Ferraro, 2002). Human parsers build up connections between words with the help of their real-world knowledge. As long as the available semantic pairs convey "good enough" meanings for understanding, parsers rapidly take advantage of that for comprehension. Here, we compared semantic influences for hierarchical and linear constructions. The inaccuracies found for even easy recursive patterns underline that good enough considerations strongly rely on semantic analyses.

Theories studying positional effects typically treat linear RB recursion as the simple "baseline" for comparisons with other more complex varieties of recursion (Christiansen & Chater, 1999). Accordingly, RB structures are argued to be processed without any difficulty (Church, 1982; Gibson, 1998; Marcus, 1980). The present results give a new perspective on what makes recursion difficult or not. For example, RB sentences were no longer simple to process when the semantic cue was inverse to the positional cue. When there was no semantic cue, RB linear sentences with only one clause were not always accurately interpreted. Our SMR suggests that not RB (as opposed to CE) is the easy default "baseline" form of recursion for the language user, but the situation in which the semantic association scheme and the positional scheme match. All conditions that deviate from this default situation, either because there is no semantic cue (and the parser has to resource pure abstract syntactical knowledge), or the semantic cue goes against the syntactical analysis (the parser is misguided by syntactical knowledge), cause difficulties.

One consequence of SMR is therefore that RB sentences are not more “basic” than CE sentences per se. For example, the sentence *the boy walks the dog that barks* would not be more frequent, basic or easy for language users than *the dog the boy walks barks*. But *the dog the boy walks barks* is predicted by SMR to be much more frequent and easier to process than *the boy the dog walks barks*. In the SMR view, it is this contrast reflected in differential frequencies and processing difficulty, between the “default” supporting semantic scheme versus the neutral or interfering semantic scheme in recursive sentences, which guides learning and everyday usage of these constructions. It is also this contrast that explains how general cognitive low level mechanisms, such as semantic memory and associative learning, provide powerful resources to guide learning. Indeed, a human learner might be exposed to default recursive sentences only in the early stage of learning (*the girl the dog bites cries*), and therefore get prepared to understand the deviations (e.g. sentences without semantic cue) from default in a later stage (*the girl the dog sees walks*).

To evaluate the SMR model further, various types of research are needed. For example, we need to know how much language users are actually exposed to semantically supported hierarchical structures and to the other types of semantic matching patterns. If, as we hypothesize within SMR, neutral and semantically mismatching sentences are largely outnumbered by semantically supporting ones, within the set of hierarchical sentences a language user comes across, this would speak for the SMR model. Notice that an analysis of the occurrence of the different types of semantic–syntactical congruency in sentences requires more than an analysis of isolated sentences of a corpus. Indeed, the semantic plausibility of a pattern of relations in a CE sentence depends on contextual factors, like discourse context, but also of the personal background knowledge of the listener. Referring to the example above, a sentence like *the girl the dog bites cries* might be easy to parse because of its description of an actual scene in the real world, but it might also be hard to parse, in the absence of such a scene, or if it is inconsistent with what happens around.

Positioning our study in the research on the learnability of hierarchical structures, our results support the low level mechanisms explanation of how humans deal with the long distances involved in hierarchical structures. In line with statistical learning models of language learning, SMR is “usage based” (Christiansen & Chater, 1999). In contrast to statistical approaches, however, the focus of our explanation for the handling of long distance dependencies is not on mechanisms that overcome positional distances (like transitional probabilities over more than one predicting element, or changes in variability of elements in given positions) (Gomez, 2002). It is the SMR concept of semantical “distance” between elements, which explains the present new data on how we deal with recursive complex linguistic constructions. In particular, why we do easily understand the hierarchical *the dog that the man walks barks*, but struggle with the linear *the dog walks the man that barks.*

Appendix

Examples of each type of stimulus sentence used in the task (CE versus RB; Matching, Mismatching and Neutral semantic-syntactic subject (A) - verb (B) relations; and 1- and 2-LoE). AiBi pairs with the same index have a syntactical subject-verb relation according to their position in the sentence. The English translations are word-by-word translations.

Sentence structure	Semantic-Syntactic relation type	LoE	Example “Kees weet dat …[1] *“Kees knows that…*
	Match	1	“…de dokter de patiënt die kermt onderzoekt.” A1 A2 B2 B1 *”..the doctor the patient who groans examines.* *A1 A2 B2 B1*

---

[1] The purpose of an introductory phrase in our materials was to disambiguate the thematic role of the second noun A2 in CE sentences. With the introductory phrase, A2 is always subject of B2 only (sentence (a)). In sentence (b), without introductory phrase, A2 can be subject of both B1 and B2. Using the introductory phrase reduced the number of syntactically ambiguous SV relations in our materials, especially the CE sentences, and allowed us to improve our measurement of accurate sentence comprehension.

(a) “*Keest ziet dat* de vader het meisje dat schreeuwt, ziet.”
A_1 A_2 B_2 B_1

(b) “Het meisje dat de vader ziet, schreeuwt.”
A_1 A_2 B_2 B_1

CE		2	“…de politie de vrouw die de hond die poept uitlaat bekeurt.”[2] A1 A2 A3 B3 B2 B1 *“…the policeman the woman who the dog that poops, walks* *A1 A2 A3 B3 B2* *arrests.* *B1*
	Mismatch	1	“…de hond de man die blaft bijt.” A1 A2 B2 B1 *”.. the dog the man who barks bites.* *A1 A2 B2 B1*
		2	“…de bouwvakker de vrouw die de auto die ronkt nafluit A1 A2 A3 B3 B2 bestuurt” B1

---

[2] In the CE sentences with 2-LoE and an introductory phrase ”Kees knew that….” used here, the *third* noun A_3, however, *could* either be subject of A_3 only, or it could be subject of both A_3 and A_2. This ambiguity is solved only if the numbers of A_2 and A_3 differ. Since the number of the nouns was kept constant in all stimulus sentences (singular, to avoid such differences in number to serve as semantic cues also), this syntactical ambiguity was present in all our CE sentences with 2-LoE. Therefore, test sentences of type A_2B_2 and A_3B_2 could in principle not be rated incorrectly, since A_2 being subject of B_2 and A_3 being subject of B_2 are both correct analyses of the sentence. To avoid ambiguity in the participants’ responses, no test sentences with A_2B_2 or A_3B_2 were used for CE sentences with 2-LoE.

		“…the worker who the woman who the car that throbs hails *A1 A2 A3 B3 B2* *drives” .* *B1*
Neutral	1	“…de jongen de vriend die valt helpt.” A1 A2 B2 B1 *“…the boy the friend who falls helps”.* *A1 A2 B2 B1*
	2	“…de vader het meisje dat de jongen die valt ziet volgt.” A1 A2 A3 B3 B2 B1 *“…the father the girl the boy who falls sees follows.* *A1 A2 A3 B3 B2 B1*
Match	1	“…de bakker het brood bakt dat rijst.”[3] A1 A2 B1 B2

---

[3] Another feature of Dutch related to embedded sentences with an introductory phrase like “Kees knew that…” used here is that the clauses are verb final. Therefore, the object in the relative clause precedes the verb (SOV), in contrast to English where the sequence within a relative clause is SVO. In RB sentences with multiple clauses, this results in SV pairs (AB) that are not adjacent, but separated by the object that is also subject of the next clause. In our materials this results in sequences with related AB pairs being separated by one other word. Our RB sentences with 1-LoE would have sequence $A_1A_2B_1B_2$, and sentences with 2-LoE $A_1A_2B_1B_2A_3B_3$. As a result, the AB (Subject Verb) pairs in the RB sentences were either separated by one element, or they were adjacent. To keep all sentences as similar as possible, we used the same introductory phrase for every stimulus sentence, at the cost of this disadvantage for RB sentences. Both the maximum distance between A and B’s in CE sentences and the mean distance were higher in CE than in RB sentences.

RB			“…*the baker the bread bakes that rises*”. *A1 A2 B1 B2*
		2	“…de groenteman de klant helpt die vraagt om de bananen die A1 A2 B1 B2 A3 rijp zijn. B3 “…*the greengrocers the customer helps who asks for banana’s* *A1 A2 B1 B2 A3* *that are ripe* *B3* .
	Mismatch	1	“…de baby de moeder troost die huilt.” A1 A2 B1 B2 “… *the baby the mother comforts who cries.* *A1 A2 B1 B2*

		2	“…de muziek de DJ aanzet die klinkt in de zaal A1 A2 B1 B2 A3 die groot lijkt.” B3 *“…the music the DJ turns on that echoes in the hall* *A1 A2 B1 B2 A3* *that looks big.* *B3*
	Neutral	1	“…het kind de oma omhelst die puzzelt.” A1 A2 B1 B2 *“…the child the grandmother hugs who puzzles”*. *A1 A2 B2 B1*
		2	“…de pastoor de man begroet die zwaait naar de bakker die A1 A2 B1 B2 A3

			fietst.” B3 *” …the priest the man greets who waves at the baker* *A1 A2 B1 B2 A3* *who cycles.* *B3*

Footnote

[1] In the Dutch translation of the RB sentence (2) slight positional changes would occur, because the object can move in front of the verb in the relative clause (being "verb final"). However, the typical contrast between short distances for the linear RB constructions and the long distances in the CE constructions are conserved in Dutch. Also, the RB clauses are lined up in a linear sequence, over time, in both languages. See also Appendix. The Dutch translation of sentence (2) and the word-by-word translation back in English are:

(2) Jan zag dat [de man een hond uitliet] [die een kat achtervolgde] [die wegrende].

John saw that [the man a dog walked] [that a cat chased] [that ran away].

A_1 A_2 B_1 A_3 B_2 B_3

[2] However, for CE stimulus sentences with 2-LoE, A_2B_2 and A_3B_2 test sentences were excluded. This is because in the Dutch sentences used, the subject of B_2 is ambiguous. Grammatically, it can be either A_2 or A_3. See Appendix.

[3] The purpose of an introductory phrase in our materials was to disambiguate the thematic role of the second noun A_2 in CE sentences. With the introductory phrase, A_2 is always subject of B_2 only (sentence (a)). In sentence (b), without introductory phrase, A_2 can be subject of both B_1 and B_2. Using the introductory phrase reduced the number of syntactically ambiguous SV relations in our materials, especially the CE sentences, and allowed us to improve our measurement of accurate sentence comprehension.

(a) "*Keest ziet dat* de vader het meisje dat schreeuwt, ziet."

A_1 A_2 B_2 B_1

(b) "Het meisje dat de vader ziet, schreeuwt."

A_1 A_2 B_2 B_1

[4] In the CE sentences with 2-LoE and an introductory phrase "Kees knew that…." used here, the *third* noun A_3, however, *could* either be subject of A_3 only, or it could be subject of both A_3 and A_2. This ambiguity is solved only if the numbers of A_2 and A_3 differ.

Since the number of the nouns was kept constant in all stimulus sentences (singular, to avoid such differences in number to serve as semantic cues also), this syntactical ambiguity was present in all our CE sentences with 2-LoE. Therefore, test sentences of type A_2B_2 and A_3B_2 could in principle not be rated incorrectly, since A_2 being subject of B_2 and A_3 being subject of B_2 are both correct analyses of the sentence. To avoid ambiguity in the participants' responses, no test sentences with A_2B_2 or A_3B_2 were used for CE sentences with 2-LoE.

5 Another feature of Dutch related to embedded sentences with an introductory phrase like "Kees knew that…" used here is that the clauses are verb final. Therefore, the object in the relative clause precedes the verb (SOV), in contrast to English where the sequence within a relative clause is SVO. In RB sentences with multiple clauses, this results in SV pairs (AB) that are not adjacent, but separated by the object that is also subject of the next clause. In our materials this results in sequences with related AB pairs being separated by one other word. Our RB sentences with 1-LoE would have sequence $A_1A_2B_1B_2$, and sentences with 2-LoE $A_1A_2B_1B_2A_3B_3$. As a result, the AB (Subject Verb) pairs in the RB sentences were either separated by one element, or they were adjacent. To keep all sentences as similar as possible, we used the same introductory phrase for every stimulus sentence, at the cost of this disadvantage for RB sentences. Both the maximum distance between A and B's in CE sentences and the mean distance were higher in CE than in RB sentences.

Chapter 5

Under What Conditions Can Recursion Be Learned?

Effects of Starting Small in Artificial Grammar Learning of Recursive Structure

---

This chapter is based on: Poletiek, F., Conway, C.M., Ellefson, M.R., Lai, J., & M.H. Christiansen (under review).

Abstract

It has been suggested that external and/or internal limitations paradoxically may lead to superior learning, i.e., the concepts of *starting small* and *less is more* (Elman, 1993; Newport, 1990). In this paper, we explore the type of structure and the type of starting small ordering that might crucially help learning. We report three artificial grammar learning experiments with human participants. In Experiments 1 and 2 we found a beneficial effect of starting small using two types of simple recursive grammars: right-branching and center-embedding, with recursive embedded clauses in fixed positions and fixed length. In Experiment 3, we used a more natural and complex center-embedded grammar with recursive loops in variable positions, producing strings of variable length. The results suggest that starting small confers an advantage for learning complex recursive center-embedded structures when the input set is organized according to structural complexity, requiring increasing computational load, but not when it is organized according to length, requiring increasing memory load.

Intuitively, learners should acquire information better when they are unhindered by internal or external limitations, such as those relating to constraints on memory or input. However, some proposals take the somewhat paradoxical stance that cognitive limitations and/or reduced input may confer a computational advantage for learning. These theories, specifically the notion that *less is more* (Newport, 1990) and the importance of *starting small* (Elman, 1993), often are couched in terms of language acquisition. When learning requires discovering relationships between component elements, as is the case in language acquisition, limited processing may be advantageous because it acts as a filter to reduce memory load as well as the complexity of the problem space, making learning more manageable. The demonstration of starting small is of central importance to both the fields of linguistics and developmental psychology, because it counter-intuitively suggests that starting with a simple initial state and limited memory capacity may make it feasible to learn complex input relationships, such as those found in language, without having to resort to innate linguistic knowledge.

Unfortunately, the evidence related to starting small is far from conclusive. Children appear to learn language better than adults; however, this result may be due to any number of factors (e.g., Hakuta, Bialystok, & Wiley, 2003). Initially, computational work supported the theory of starting small (e.g., Elman, 1993), but more recent simulations appear to contradict those findings (Rohde & Plaut, 1999, 2003). Further, empirical data gathered from human participants have not resolved the issue; some data support starting small, (Cochran, McDonald, & Parault, 1999; Kareev, Lieberman, & Lev, 1997; Kersten & Earles, 2001; Lai & Poletiek, 2011; Poletiek, 2011), while other data do not (Fletcher, Maybery, & Bennett, 2000; Ludden & Gupta, 2000; for reviews see Rohde & Plaut, 1999, 2003).

This paper seeks to determine under what conditions, if any, starting small might have an effect on learning complex recursive language-like structure. We investigate the limits of the *starting small* hypothesis using the artificial grammar learning (AGL) paradigm. First, we discuss the inconclusive evidence for starting small and two possible explanations of the effect (structural complexity versus memory load). Second, we present three experiments to examine the starting small effect using recursive artificial grammars.

Experiment 1 shows that when the input of a simple right-branching recursive grammar is staged according to the number of recursive loops at the end of strings, participants achieve better learning than when the input is randomly ordered. Experiment 2 shows that this facilitation also occurs for a more complex center-embedded grammar. Experiment 3 directly compares the effect of starting small according to structural complexity versus item length. The results of Experiment 3 replicate the facilitation of starting small for the center-embedded grammar, and more importantly reveal that the starting small effect is due to the structural characteristics of the grammar becoming more salient rather than due to changes to item length. Based on these findings, we propose that the facilitation effect of reduced input occurs for both simple recursive structures and complex ones, but only when the input 'grows' according to structural complexity and not according to increasing item length. In sum, these findings point to a fundamental influence on learning that has far-reaching consequences for language acquisition, development, and inductive learning more generally.

Starting Small Evidence

The *less is more* and *starting small* hypotheses can be thought of as two related but separate ideas. The ideas are similar in that they propose that processing limitations may present a learning advantage, but they differ in terms of the nature of the limitation itself. Processing limitations may arise from internal cognitive constraints, or from external constraints, in the form of staged or incremental input. Orthogonal to this distinction, external or internal limitations may apply either to the volume or to the complexity of the information. As a result, the cognitive *less is more* hypothesis may refer to the benefit of internal limited memory capacity or to computational capacity (though both cognitive functions may be related, Baddeley, 2000). Analogously, in the external version of the hypothesis, *starting small*, may refer to the benefit of the limited amount of information in the input items (e.g., length) or to their limited structural complexity. Here, we review data related to all these possibilities, starting with the internal/cognitive version.

In the context of language acquisition, Newport (1990) proposed that maturational constraints in the form of cognitive limitations are crucial for allowing language to be learned successfully. In support, data were reported from deaf adult participants, who learned American Sign Language (ASL) at different ages. On ASL morphology and syntax, native signers outscored early learners, who in turn outscored late learners. Newport suggested that young children have to focus on the smaller segments of language -- where smaller segments refer to smaller amounts of information -- because of their limited working memory capacity. The children become proficient with the constituent parts of signs first, and then learn to combine them into larger, more complex signs. Late learners, because they do not have the same cognitive limitations, attempt to learn larger and complex wholes in their entirety. Although the late learners learn quickly compared to the early learners, they are less proficient at combining simple constituents into more complex wholes.

In a subsequent study exploring the *less is more* hypothesis (using the related term, “starting small”), Elman (1993) trained a simple recurrent network (SRN) to learn aspects of an artificial language. Under standard conditions, the network was unable to learn the sequential regularities of the grammar. But when Elman simulated children's working memory limitations by periodically eliminating the network's access to its prior internal states—and allowing the size of this temporal window to increase over time—the neural network's performance improved.

Further support for the *less is more* hypothesis comes from Cochran, McDonald, and Parault (1999) who taught adults portions of a modified version of ASL. They simulated cognitive computational limitations by supplying a simultaneous capacity-limiting task during training and found that the participants in the no-load condition displayed more rigid learning and were less adept at using the signs in new contexts. Additionally, Kareev, Lieberman, and Lev (1997) explored the relation between working memory capacity and the detection of correlation. Human participants were tested on their ability to predict the relationship between two binary variables. Participants with lower working memory capacity were better at detecting the appropriate correlation and performed better on the task than did high working memory capacity participants. Since working memory, on this account, has both a short term storage and a computational cognitive function, this evidence was taken to support the hypothesis that less cognitive capacity confers an advantage in some inductive learning tasks.

However, there may be reasons to be critical of these data. For instance, Rohde and Plaut (1999, 2003) conducted neural network simulations that contradicted Elman's (1993) results. Using the same architecture, simulation parameters, and training input, Rohde and Plaut failed to get an advantage for reduced cognitive capacity. They also questioned a number of previous conclusions (Cochran et al., 1999; Kareev et al., 1997), instead arguing that these earlier data do not support the notion that internal limitations benefit learning. Other studies appear to support this perspective. For example, adult participants in an AGL task with a capacity-limiting condition failed to show an effect of starting small (Ludden & Gupta, 2000). In a similar vein, younger children do not surpass older children in an implicit covariation detection task (e.g., Fletcher et al., 2000).

Whereas these studies reviewed so far focus mainly on the internal version of less is more – that is, whether limitations on memory capacity result in learning benefits -- there are fewer experiments testing the external constraints version of the hypothesis. The lack of research exploring whether limiting or staging the input confers learning advantages may be partly because of the widespread belief that the language input children receive is not substantially different from adults. However, as Rohde and Plaut (2003) point out, there is evidence that child-directed speech tends to consist of shorter utterances and less complex sentences than adult-directed speech (e.g., Pine, 1994; Tomasello, 2003), requiring less memory and computational load to process. Therefore, it may be feasible that starting with simplified and shorter utterances provides a learning advantage, and that this may help explain children's efficiency in acquiring natural language. Elman (1993) and Rhode and Plaut (1999) provided a test of this version of starting small using neural network simulations. The results are mixed. In an incremental input condition, Elman organized the network's input so that it was exposed only to simple and short sequences first. Afterwards, complex sequences were introduced to the network gradually. The grammar used by Elman had recursive rules generating center-embedded exemplars. Starting small was implemented by presenting the network with exemplars having increasing numbers of levels of embedding. When trained in this way the networks showed a learning advantage; however, Rhode and Plaut (1999) did not replicate this starting small effect in a similar computer simulation.

A few recent studies with human participants seem to support the validity of an external constraints view of starting small (Kersten & Earles, 2001; Lai & Poletiek, 2011; Lany, Gomez, & Gerken, 2007). Kersten and Earles (2001) exposed adults to an artificial language comprising both auditory nonsense sentences and visual, animated events. Some of the participants were exposed to a staged input regimen, in which they received input in three phases: first only single words were presented along with the animated events, then sentences composed of two words, then finally three-word sentences. These participants fared better on tests of their understanding of the language compared to participants who were exposed to a non-staged random input presentation. Though Kersten and Earles view this demonstration as supporting the notion of internal limitations providing a starting small advantage, Rohde and Plaut (2003) note that these data show the possible benefits of using a staged input training scheme. Likewise, in the study by Lany, Gomez and Gerken (2007) participants only acquired a complex acXbcY language in which the co-occurring aX and bY were separated by a varying c- element when they were first trained with a simple version of the language without the c-element, i.e. the aXbY structure. This result is in line with an external starting small effect. Finally, Lai and Poletiek's (2011) study replicated the beneficial effect of a starting small regimen found by Elman (1993) using an artificial center-embedded grammar that gradually increased in complexity. Though Lai and Poletiek found a strong facilitation of starting small, the center-embedded pattern they used was quite simple as compared to Elman's natural stimuli. Moreover, superficial phonological cues in addition to the recursive center-embedded structure provided information about the underlying recursive dependencies in the grammar, likely making it easier to learn than center-embedded constructions without such additional cues.

To sum up, we note three crucial observations. First, a number of empirical studies suggest that internal cognitive constraints seem to provide an advantage for learning, although the computer simulation studies are more inconclusive in this regard. Moreover, since memory and computation are closely related cognitive functions, it is still unclear which of these two aspects – memory vs. computational load -- is responsible for the learning advantage. Second, a few studies show that external constraints – i.e., limited complexity or quantity of information in the input - may enhance learning as well. However, from the studies on the external starting small effect (Kersten & Earles, 2001; Lai & Poletiek, 2011), it is also unclear which type of limitation placed on the stimuli (reduced complexity or reduced length) crucially affects the learnability of the underlying structure, because manipulations of stimulus complexity often co-vary with stimulus length. Third, it is possible that the type of structures used to test starting small may affect the outcome. Complex structures may be hierarchically recursive, with long distance dependencies, like center–embedded structures. Alternately, recursive constructions may not be hierarchical but linear, adding recursive clauses at the end of strings, as in right branching recursion. Simple, finite state grammars may also be presented in a starting small fashion. One of the major successful tests of starting small incorporated a complex hierarchical natural recursive structure (Elman, 1993), whereas one of the "unsuccessful" tests used a simple standard finite state grammar (e.g., Ludden & Gupta, 2000, Experiment II). Thus, it is possible that the advantage of starting small depends partly on the underlying structure to be learned.

Here, we explore the possibility that starting small may facilitate the learning of recursive constructions specifically. We suggest that the learning of a particular recursive structure involves two parts: a) learning the structural regularities defining the construction in its base (non-recursive) form, and b) learning to generalize these regularities in a recursive manner. Starting small allows for the separation and subsequent learning of these two parts, by displaying only the basic regularities in the first stage of exposure, and the recursive generalizations in later stages, after the basic structure has been mastered (see also Poletiek, 2011). Because mastery of the basic regularities is *key* to successful processing of sequences with recursive embeddings, the time course of the learning process is crucial. Hence, presenting the input in a starting small fashion with additional recursive generalizations at each subsequent stage, may optimally support this learning procedure. This possibility is particularly interesting in the light of the recent ongoing debate about the cognitive mechanisms supporting the acquisition of recursion in natural language, and the role of the stimulus input in this learning process (Chomsky, 1995; Christiansen & Chater, 1999; Christiansen & MacDonald, 2009; Corballis, 2007; de Vries, Christiansen, &

Petersson, 2011; Fitch & Hauser, 2004; Gibson & Thomas, 1999; Perfors, Tenenbaum, & Regier, 2011)

Based on these observations, we explore starting small experimentally, proposing that recursion learning by induction is helped if the input is organized in a starting small fashion. For simple right branching structures, we propose that both types of constraints (on memory load and on computational capacity) help learning. As the structure gets increasingly complex and computationally more demanding, (e.g., hierarchical with long distance dependencies), limits on stimulus complexity but not on memory load will effectively enhance learning. Before presenting the three experiments that explore this hypothesis, we briefly describe the types of recursive grammars used in the present methodology.

Recursive Artificial Grammars

A recursive grammatical construction is one that is defined by self-reference. Different types of recursion can be found across a variety of natural linguistic structures. As the amount of self-referencing increases within a recursive construction, the amount of embedding increases. Consider the grammatical English noun-phrases in (1):

1. a) *The dog* [*on the sidewalk*].
   b) *The dog* [*on the sidewalk*] [*near the tree*].
   c) *The dog* [*on the sidewalk*] [*near the tree*] [*by the house*].

The above sentences involve *right-branching* recursion, in which new prepositional phrases are recursively added onto the right end, creating sentences of potentially infinite length. Sentence (1a) comprises 0 level of embedding (LoE), (1b) 1-LoE, and (1c) 2-LoE.

Increased levels of right-branching embedding result in slightly decreased comprehensibility of English sentences (Christiansen & MacDonald, 2009). Decreases in comprehension are even larger for a second type of recursive structure: *center-embedding* (e.g., Bach, Brown, & Marslen-Wilson, 1986). Center-embedded recursion grows a sequence by embedding new material in the center, and pulling apart elements that depend on each other, resulting in a hierarchically built up string having long distance dependencies. For example, consider the sentences in (2):

2. a) [*The boy likes the dog]*.

b) [*The boy* [*the girl loves*] *likes the dog]*.

c) [*The boy* [*the girl* [*the woman admires*] *loves*] *likes the dog]*.

As before, sentence (2a) comprises 0-LoE, (2b) 1-LoE, and (2c) LoE.

The same semantic relationships can be expressed using either right-branching or center-embedding recursion. For example, consider the two sentences (without recursive embeddings) having the same basic structure (3a and 3b), below. These two sentences can be combined either using right-branching embedding as in (3c) or center-embedding as in (3d):

3. a) [*The boy likes the dog*].

b) [*The girl loves the boy*].

c) [*The girl loves the boy*] [*who likes the dog*].

d) [*The boy* [*whom the girl loves*] *likes the dog*].

Both sentences express similar semantic content and involve equal lengthening of the sequence, though the center-embedding construction is presumably more complex than the right branching version because it involves long distance dependencies. Thus, whereas both sentences (3c) and (3d) appear to involve equivalent memory load (due to equal lengths of the sentences), they appear to differ in terms of computational complexity. Thus, by comparing performance on right-branching and center-embedded stimuli, it may be possible to experimentally disentangle the factors of memory load versus computational complexity in starting small.

Translating this comparison into a controlled experimental situation, we first constructed a right branching (Experiment 1) and a matched center-embedding grammar (Experiment 2), to test the effect of a starting small exposure (i.e., gradually increasing the number of embeddings in the input) on learning these two types of grammars. In Experiment 3, we more directly explored the separate contributions of constraints on memory load versus constraints on computational load in terms of leading to a learning advantage.

To generate letter sequences used in our first two experiments, we created two categories of letters: Category A and Category B. Category A letters could be paired to

Category B letters. The first letter from the pair belonged to Category A and the second letter of the pair belonged to Category B. Furthermore, we included two subsets within each category: Subset 1 and Subset 2. Translating this grouping in natural language syntactical categories, Category A elements might be thought of as nouns, and Category B letters as verbs. Moreover, Subset 1 and Subset 2 elements might represent singular and plural items, respectively. Accordingly, letters from Category A - Subset 1 could be paired only with letters from Category B – Subset 1. Similarly, letters from Category A – Subset 2 could be paired only with letters from Category B – Subset 2. Twelve consonants, C, Q, M, P, X, S, W, Z, K, H, T, and V represented the subsets within each category. A recursive rule was used to generate self-embedded exemplars. The embeddings were either right branching (added at the end of the exemplar) or center-embedded (inserted in the exemplar), depending on the type of grammar, as indicated in Figure 1a and Figure 1b respectively.

In Figure 1a, one of the two paths starting from S3 represents the recursive loop generating a right branching clause. The other path from S3 terminates the string. In Figure 1b, one of the two paths from S1 and S2 represents the recursive loop generating a center-embedded clause. For an example of how these grammars generate recursive input strings, C[PH]W was produced from the grammar of Figure 1b, having one level of center embedding. CW[PH][QZ] was produced from the grammar of Figure 1a, having two levels of right branching embedding.

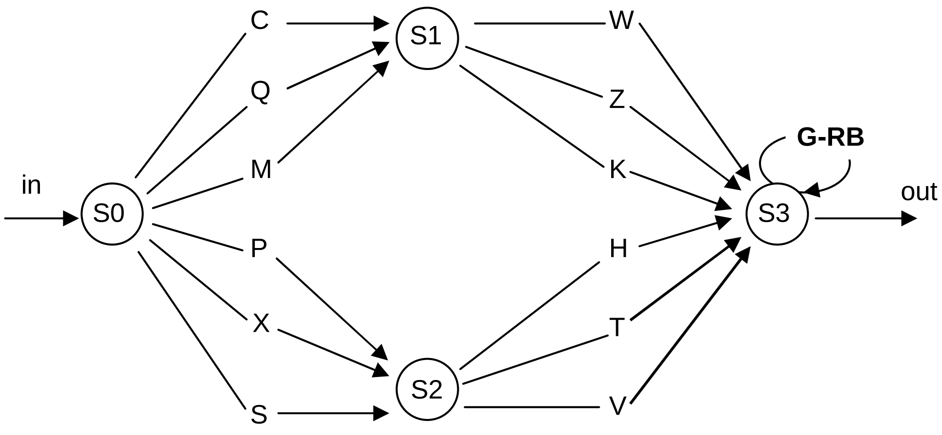


Figure 1a: Right Branching Grammar G-RB used in Experiment 1.

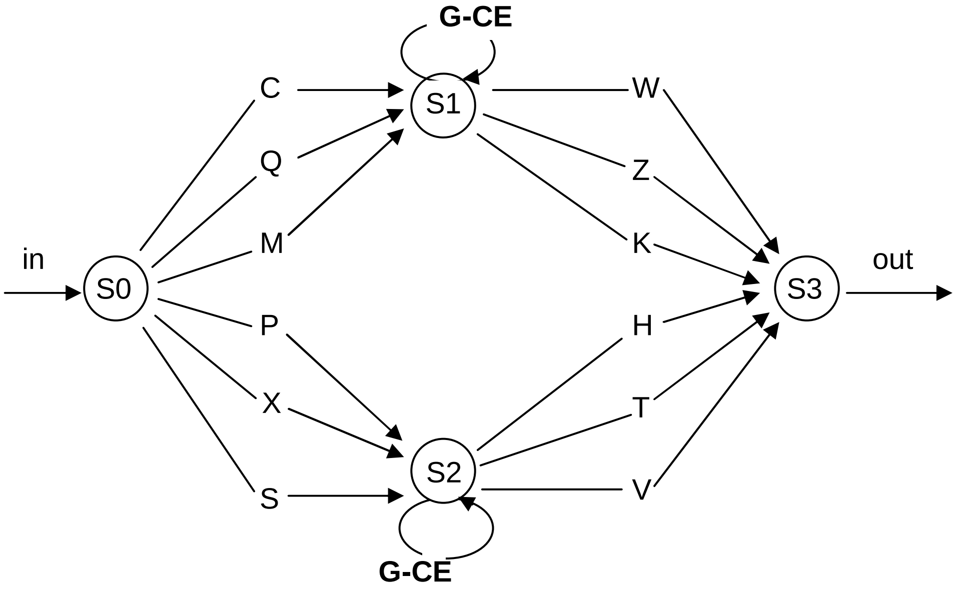


Figure 1b: Center embedding Grammar G-CE used in Experiment 2.

To first assess the hypothesis that starting small helps with learning right-branching recursion, we conducted Experiment 1, testing whether a simple recursive grammar will produce the starting small effect when exemplars are ordered according to increasing numbers of levels of embedding and string length.

Experiment 1

In the first experiment, we generated letter strings from an artificial grammar having right-branching recursion (Figure 1a). We ordered the exemplars differently for two separate training conditions. In the starting small condition, exemplars were ordered according to increasing levels of embedding (LoE). This corresponded to first presenting strings with 0-LoE, then strings with 1-LoE, and finally strings with 2-LoE. In this way, the input “started small" with basic sequences only and progressively became more complex with applications of the right branching rule. In the second training condition, participants received the same input but presented in random order. We predicted that by ordering the strings in this way, the starting small input group would learn the basic structure of the input first and then be able to generalize it to more complex recursive structures, providing an advantage over the random group, which is exposed to both the basic and the recursive constructions in an intermixed fashion.

Method

Participants. For Experiment 1, 14 undergraduate participants (seven in each condition) were recruited from Psychology classes at Cornell University, earning extra credit.

Materials. The stimuli were letter sequences generated from the artificial grammar displayed in Figure 1a (see Appendix A). The sequences were based on the repetition of pairs, within a recursive structure, in which arbitrary letters assigned to Subset 1 and Subset 2, and to Category A and Category B (see Figure 1a). An example of a 0-LoE sequence is CW, a 1-LoE sequence is CWPT, and a 2-LoE sequence is CWPTQZ.

Unique sequences were created for the training and test sessions. Fifty sequences comprised the training session. Of these 50 training sequences, 10 were 0-LoE embedding, 20 were 1-LoE embedding, and 20 were 2LoE. An additional fifty sequences comprised the test session (see Appendix A). Of these test sequences, 25 were generated from the same grammar as the training sequences (grammatical) and 25 did not follow the grammar (ungrammatical). Ungrammatical sequences were created by changing one letter of a grammatical test sequence. The substituted letter was one that was of the proper noun-verb category but with an incorrect plurality (subset). The positions in which the substituted letters occurred in the sequences were distributed evenly across all items. The test session comprised 16 sequences of 0-LoE, 16 of 1-LoE, and 18 of 2-LoE, with each level of embedding having half grammatical and half ungrammatical structures.

Procedure. The experiments were run using the E-Prime presentation software with stimuli presented on a computer monitor. Participants were randomly assigned to one of two conditions: Starting Small or Random. All participants were instructed that they were participating in a memory experiment. They were told that in the first part of the experiment they would see sequences of letters displayed on the screen and that they would be tested later on what they observed. Each sequence in its entirety was presented individually, for a duration of four seconds each. Each of the 50 training items was presented three times, for a total of 150 input exposures. The starting small participants received staged input: three blocks of the 0-LoE sequences were presented first; next three blocks of the 1-LoE sequences, and finally three blocks of the 2-LoE sequences. Sequences were randomized within blocks. The random group received all the sequences across all LoE intermixed with one another, in random order. Thus, both the starting small and the random groups received the same training input but in different orders of presentation.

After the training phase, participants were told that the items they had just seen had been generated by a complex set of rules that determined the order of the letters. They were instructed that they would now see new letter strings, some of which followed the rules of the grammar, and some of which did not. Their task was to classify whether each letter string followed the same rules as the training sequences or not, by pressing a button marked “YES" or “NO". Both the starting small and random groups received the same test instructions and the same set of 50 test sequences were presented in random order for each participant.

Results and Discussion

The mean percent correct classification of the 50 test items was 70.0% for the starting small group (*M* = 35.00, *SD* = 3.79) and 54.9% for the random group (*M* = 27.43, *SD* = 4.79). We conducted single group *t*-tests and found that only the starting small group performed significantly above chance levels (t (6) = 6.99, $p < .001$). The starting small group also performed significantly better than the random group (t (6) = 3.86, $p < .01$).

The results of Experiment 1 show that only when the input was presented in a staged fashion, with 0-LoE strings presented first, were participants able to successfully learn the right-branching recursive structure of the artificial grammar. The recursive structure was not learnable when the training items were presented in random order. Crucially, the starting small group out-performed the random group, lending empirical support to the starting small hypothesis.

Experiment 2

In Experiment 1, we observed an effect of starting small for a relatively simple recursive grammar. Right-branching recursion here involves the addition of new basic 0-LoE structures at the end of a grammatical sequence. In the resulting grammatical sequence, the grammatical dependencies are all between adjacent elements in a string. We next explore to what extent the starting small effect is also present in the more complex and computationally demanding center-embedding recursion, which is characterized by *non-adjacent* dependencies (Figure 1b); here, the basic 0-LoE structure has to be recognized even if the two connected elements it is made of (an A category and a B category letter) are pulled apart to distant positions.

We predicted that by ordering the strings, the starting small input group would be able to generalize the basic agreement structure from the 0-LoE items to the more complex center-embedded constructions. In contrast, the random group was expected to have problems learning this grammar as they were presented with both basic and recursive constructions intermixed with one another. However, as the center-embedded operation is more complex, lower performance is expected than for the right branching structure, when participants are provided with the same number of learning items as in Experiment 1.

Method

Participants. For Experiment 2, 16 undergraduate participants (eight in each condition) were recruited from Psychology classes at Cornell University, earning extra credit.

Materials. The sequences used in Experiment 2 were identical to those in Experiment 1 except that they were converted from a right-branching to a center-embedded structure (see Appendix B). That is, embedding was increased by inserting additional noun-verb pairs into the middle of the center-embedded sequences to achieve higher levels of embedding. An example of a 0-LoE center-embedded sequence is CW, a 1-LoE sequence is CPTW, and a 2-LoE sequence is CPQZTW.

Procedure. The procedure was identical to Experiment 1.

Results and Discussion

The mean percent correct classification on the 50 test items was 63.0% for the starting small group (*M* = 31.5, *SD* = 4.71) and 52.8% for the random group (*M* = 26.4, *SD* = 1.06). Only the starting small group performed significantly above chance levels (t (7) = 4.08, $p < .005$). The starting small group also performed significantly better than the random group (t (7) = 2.88, $p < .05$). Similar to Experiment 1, the results show that only when the input was presented in a staged fashion were participants able to successfully learn aspects of the recursive structure of the artificial grammar. Thus, Experiment 2 replicates the starting small effect and extends its applicability to the more complex center-embedded structures. Furthermore, although both Experiments 1 and 2 showed a facilitative effect of starting small, the limitations of the computational load entailed by the non-adjacent nature of the center-embedded grammar may explain the higher performance in Experiment 1 as compared to Experiment 2.

Experiments 1 and 2 used recursive grammars with the same basic structure, pairing two elements of two categories A and B, but having different recursive operations. Given that all pairings had equal lengths (i.e. an A with a B letter), strings with an equal number of embeddings necessarily have equal lengths in both grammars G-RB and G-CE: 0-LoE strings have two letters, 1-LoE strings have four letters and 2-LoE strings have six letters. As a result, the starting small ordering according to number of LoE's correlates perfectly with ordering according to increasing length. Therefore, the results of Experiments 1 and 2 are inconclusive with respect to the relative contributions of memory load (via manipulation of input length) as compared to structural complexity (via manipulation of LoE). Though string length has been suggested to affect learning independently from complexity in a non-recursive AGL study (Poletiek & van Schijndel, 2009) previous findings on the *less is more* and *starting small* effects with recursive grammars have not distinguished between these two contributions.

From the perspective of reducing memory load, it may be that staging the training input according to increasing string *length* will facilitate learning. On the other hand, the alternative view is that reduced *complexity* at the beginning of learning – i.e., gradually increasing the levels of embedding over time - is the more important factor causing the starting small facilitation. As we proposed above, the learner must first master the basic structural patterns before these can be generalized to recursive constructions. The starting small procedure forces the learners to focus on the basic patterns before they encounter the increasingly more complex recursive structures. Hence, we hypothesize that starting small helps because it gradually introduces more and more complex recursive structure following the initial exposure to the basic pattern, not because it incrementally stages the *amount* or *length* of input per se.

In natural language, recursive constructions are more complex than the two-element pairings used in Experiments 1 and 2. In Experiment 3, we used a more realistic though still artificial implementation of recursion in a grammar producing strings of variable length with equal LoE's. Furthermore, a starting small training scheme according to string length is compared with a starting small training scheme according to string complexity, to determine the relative impact of these two aspects of starting small: memory and computational constraints.

In summary, our results so far have suggested that starting small occurs for recursive grammars but leaves unanswered whether the effect was caused by the gradual increase of string length or structural complexity. Notice that the answer to this question has implications for the issue of the learnability of complex center-embedded structures by exposure to exemplars. If a non-linguistic environmental cue (incremental ordering over time) in the input can be exploited effectively to learn underlying structure, this might strengthen the possibility that complex structures, like center-embedded constructions in natural language, may be learnable from environmental information in the input (see also Christiansen & Chater, 1999; Christiansen & MacDonald, 2009; Poletiek & Chater, 2006).

Experiment 3

In the same manner as in Experiment 1 and 2, a recursive center-embedded artificial grammar was used with two categories A and B, and eight letters. However, the basic elements in each category (A and B) were either individual letters or bigrams. Category A elements were C, QP, S, and Category B elements were WZ, K, V. Category A elements could be paired with category B elements from the same subset, as displayed in Figure 2.

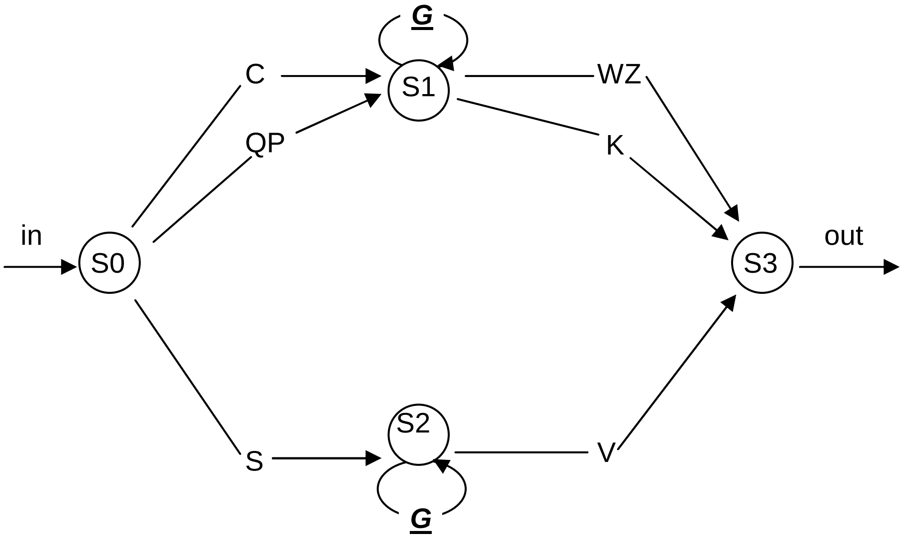


Figure 2: Center embedding grammar, G, used in Experiment 3, with exemplars varying in length for an equal number of LoE. E.g., QP[CWZ]WZ (length 7) and S[CK] V (length 4) are exemplars of this grammar having both 0-LoE.

This resulted in a grammar G with the same structural characteristics as Grammar G-CE, but having fewer legal AB pairings. As can be seen in Figure 2, five unique legal AB pairs (0-loE sequences) could be generated by G (as compared to G-CE having 18 unique legal AB pairs). In this manner, we reduced the variability of G to compensate for its increased complexity caused by the variability of string length, in order to make learning possible within the context of the experimental task.

Method

Participants. To allow a valid comparison between the two starting small regimens, in addition to comparing starting small with a random regimen, we enhanced the statistical power of our test, with increased sample size. Fifty-four students from Leiden University participated, either for course credits or financial compensation (€ 4.50).

Materials. Fifty grammatical sequences were generated from the grammar G, for the training set (Appendix C). Each exemplar was presented three times. In the structure based Starting Small (SS-S) condition, the exemplars were presented successively in three consecutive blocks of five 0-LoE sequences, followed by fifteen 1-LoE sequences, and finally thirty 2-LoE exemplars. Within a block, the ordering of the strings was randomized and each unique sequence was presented three times (Appendix C). Following the same procedure, the same fifty sequences were ordered according to their length in the Starting Small according to length (SS-L) condition. In the SS-L condition, blocks were thus determined by string length. Ten blocks were presented successively: a block of two unique strings with length 2, followed by a block of two unique strings of length 3, two strings of length 4, four strings of length 5, eleven strings of length 6, seven strings of length 7, eleven strings of length 8, eight strings of length 9, two strings of length 10 and one string of length 12 (see Appendix C).

Within a block, the sequences were presented randomly. As in the SS-S condition, within one block, the unique sequences were presented three times each in a random order, with the constraint that one unique sequence could not be repeated. For blocks with two strings, the strings were alternated three times. The string in the last block (one string of length 12) could of course not satisfy the non-repetition requirement. It was repeated three times. In the Random condition, the 50 strings were presented in random ordering in one single block, each three times. No subsequent repetitions could occur in the random presentation.

The test set was made of 25 grammatical and 25 ungrammatical strings. As in Experiment 1 and 2, ungrammatical sequences were created by changing one element of a grammatical test sequence. The substituted element was one that was of the proper category (a B was replaced with a B element) but from an incorrect subset, hence making an incorrect pair with the corresponding A element. The positions in which the substituted elements occurred in the sequences were distributed evenly across all items (Appendix C). Since the present grammar G generated only five unique 0-LoE sequences, these could occur in both the training and test set. The 1-LoE and 2-LoE test items did not occur in the training set.

Procedure. As in Experiment 1 and 2, E-Prime presentation software was used with stimuli presented on a computer monitor. Participants were randomly assigned to one of three conditions: Starting Small-Structure based (SS-S), Starting Small-length based (SS-L) or Random. All participants were instructed that they were participating in a memory experiment. In each condition, the same 50 training items were presented three times, in successive blocks, for a total of 150 input exposures. Depending on condition, blocks were determined on the basis of structure (number of levels of embedding) in the SS-S group, and on the basis of string length in the SS-L group. In the random condition, the input was randomized and presented in one block (Appendix C).

Results and Discussion

The mean number of correct classification was 61.0% for the SS-S group (*M* = 30.5, *SD* = 8.5), 51.0% for the SS-L group (M= 25.5; *SD* = 3.0) and 45.0% for the random group (*M* = 23.5, *SD* = 3.0). A one-way ANOVA indicated a significant main effect of condition (*F* (2, 51) = 9.6, *p* < .001). One-sample *t*-tests revealed that only the SS-S group performed significantly above chance levels (*t* (17) = 2.8, *p* = .012). The SS-S group performed significantly better than both the SS-L group (*t* (20.6) = 2.5, *p* < .02) and the random group (*t* (21.5) = 3.6, *p* < .001). However, performance in the SS-L and the random group did not differ.

To analyze the knowledge learned under different training conditions in more detail, especially how performance related to the levels of embedding in the test items, the accuracy of grammaticality judgments for test items with 0-, 1- and 2-LoE were calculated separately for each condition (see Figure 3). An ANOVA of judgment accuracy with Condition as a between-subjects variable, and LoE as a within-subjects variable indicated a significant main effect of the number of LoE in the test item (($F(2) = 4.3$, $p = .016$) and a significant interaction between Condition and the number of LoE ($F(4) = 3.3$, $p = .014$). As can be seen in the figure, the interaction effect was mainly due to the pattern of judgments in the random group. For the group trained with a starting small regimen according to structure, 0-LoE items were classified more accurately than 1-LoE items (Mean difference = .09, $t\,(17) = 2.5$, $p = .024$) and 2-LoE items (Mean difference = .13, $t\,(17) = 2.8$, $p = .011$). However, 1-LoE items were not judged more accurately than 2-LoE items (Mean difference = .03, $t\,(17) = 1.3$, $p > .10$). A similar pattern of performance for different levels of embedding items was observed for the starting small group according to length (Mean difference between 0- and 1-LoE items = .11, $t\,(17) = 2.4$, $p = .024$; mean difference between 0- and 2-LoE items = .10, $t\,(17) = 2.4$, $p = .024$). Notice, however, that overall performance for this group was not above chance. On the other hand, the random group did not show this same pattern of better performance on the 0-LoE items.

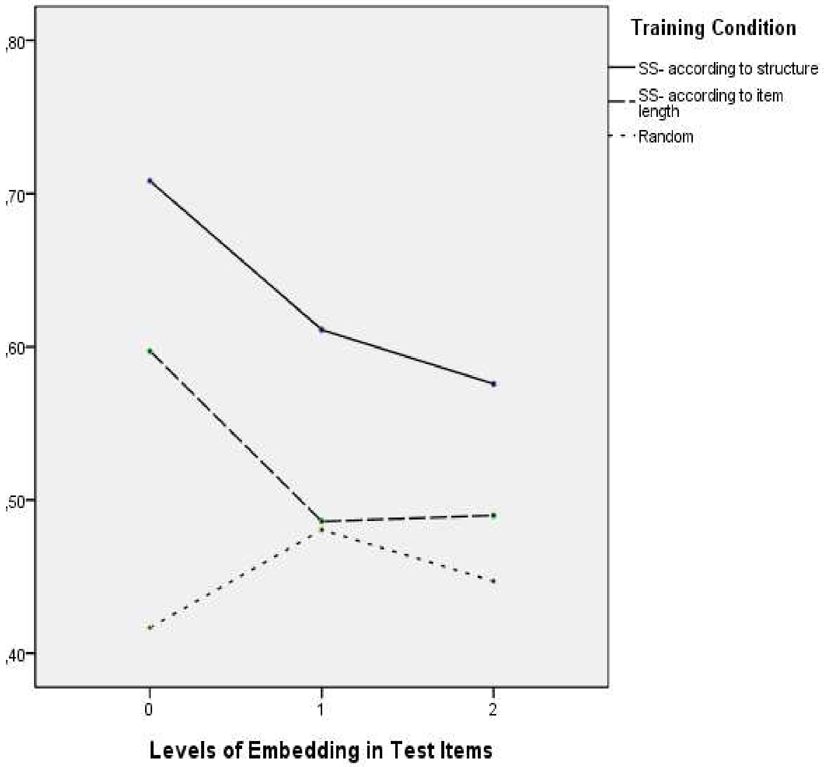


Figure 3: Performance for test items with 0-, 1- and 2-LoE, for each training condition (Random, Starting Small according to structure, and Starting Small according to item length).

In summary, the results of Experiment 3 showed facilitation of starting small for learning the center embedded grammar only when the training items were staged according to increasing LoE, and not when the items were staged according to increasing string length. For both types of starting small regimens, 0-LoE test items were judged more accurately than 1- and 2-LoE test items. However, among the complex test items with embeddings, 1-LoE items were not judged better than 2-LoE items. Strikingly, for this type of complex center embedded structure, for a random presentation, even the 0-LoE test items could not be recognized as to their grammaticality.

General Discussion

Our three experiments provide unique insight into when starting small in the form of staged input may help the learner. For three artificial recursive grammars with a self-embedding structure, a starting small presentation of the input was compared with a randomly ordered presentation. For all grammars tested - i.e., a linearly right branching recursive grammar, and two more complex center embedded recursive grammars - the starting small presentation was a necessary condition for learning. With randomly ordered presentation, no learning was demonstrated. The starting small facilitation relies on constraining two aspects of the stimulus input. First, starting small reduces the length of the sequence units to be processed initially, and second, it reduces the computational complexity of the initially encountered stimuli. Experiment 3 disentangled the effect of length from that of complexity, and showed that only by reducing complexity, does starting small help learning.

The results clearly suggest that constraining the complexity of the input effectively facilitates learning the complex self-embedding recursive structure. Participants trained with an input merely growing in string length but disregarding the structure of the stimuli showed no learning, except for the items without recursive loops. To sum up, our data suggest that artificial recursive structures varying in complexity from simple linear ‘additive’ right-branching structures to complex center-embedding constructions involving long distance dependencies at fixed or variable locations in the sequences require a staged input of exemplars to be learned. More specifically, the sequences to which the learner is exposed to should grow over time in terms of structural complexity, not merely in length, to cause an effective facilitation.

Our findings are in line with previous studies demonstrating the difficulty of learning artificial complex structures that mirror natural grammar complexity. Under such conditions, either no structure learning could be demonstrated (de Vries, Monaghan, Knecht, & Zwitserlood, 2008) or it could be demonstrated only if extra-linguistic cues in the input environment were present (Lai & Poletiek, 2011). Elman’s (1990) computational study first demonstrated the beneficial effect of starting small in a computational environment, using a grammar similar to the ones used in the present research. However, the present study is the first to investigate and compare the computational (ordering according to structure), and the memory aspects (ordering according to length) of a starting small input as compared to random ordering.

Interestingly, several recent AGL studies on the learnability of complex structure include starting small schemes in the design, even if it was not the focus of the study, which might have contributed to the positive results reported. For example, Perfors and colleagues (2011) proposed a Bayesian computational model for inductive learning of a complex artificial phrase grammar. The computational model was run with artificial input data based on features of child directed speech, with items growing according to level of complexity. Bahlmann, Schubotz and Friederici (2008) compared participants learning a non-recursive artificial grammar with a group learning a recursive artificial grammar using fMRI. Both input sets were organized in a starting small fashion. And, a recent event-related potential (ERP) study investigating the neurophysiological correlates of artificial grammar learning only could elicit learning from adult participants when the materials were presented in a starting small fashion (Christiansen, Conway, & Onnis, 2012). These studies further underscore the importance of starting small, especially for the learning of recursive constructions.

Translating these results back to the natural situation, what does it tell us about natural language learning? To answer this question, first, we need to compare the artificial grammar implemented in our study with natural language, and secondly, we need to compare the starting small procedure in the lab, with the linguistic environment of a language learner, i.e., child directed speech. In natural language, recursive constructions occur quite frequently. In most cases, self-referring recursive regularities form simple left- or right-branching structures as in repeating adjectives (*the* [*big*] [*red*] [*plastic*] *ball*) and repeating sentential complements ([*Mary says*] [*that Bob thinks*] [*that Gabby saw Bill*]), respectively. More complex self-embedded structures are much less frequent and typically limited to a single level of embedding. Sentences with two or more levels of embedding (as in *The boy* [*the girl* [*the woman admires*] *loves*] *likes the dog*) are very difficult to understand (e.g., Blaubergs & Braine, 1974; Wang, 1970—see Christiansen & MacDonald, 2009, for a review) and practically absent from spoken language (Karlsson, 2007). The better learning we observed for the right-branching structures in Experiment 1 compared to the center-embedded structures in Experiment 2 might be seen as reflecting the distributional asymmetry found in natural language between these two types of recursive constructions

Although biological factors appear to provide important limitations on the learning of self-embedded recursive structure (de Vries et al., 2011), the experience that a learner has with particular recursive constructions also play a key role (Christiansen & MacDonald, 2009; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Our results suggest that the specific order with which learners experience recursive structures may play an important role in how well such recursive constructions can be mastered. Specifically, starting small enables learners to focus on learning the basic structural patterns first before they are faced with the more complex embedded structures. Hence, if natural language input is structured in a way similar to our starting small procedure, then we would expect facilitation for learning recursive structures both in artificial grammars, as here, and in natural language, more generally.

The second comparison to assess the validity of the present result for natural language is between the two starting small procedures (structure-based and length-based) and natural child directed speech. If the constraints on computational capacity effectively enhance learning in the natural situation, then the sentences a learner is exposed to should become gradually more complex over time, rather than longer. Likewise, sentences occurring in child directed speech would be expected to be limited mainly in complexity, but not necessarily in length. But is this in fact the case?

Indeed, studies on early language acquisition consistently find that the language of primary caregivers includes fewer complex sentences, and sentences containing no or fewer subordinate clauses than adult speech (Brown, 1973; Pine, 1994; Tomasello, 2003). In addition, the structural complexity of early language input is reduced by other features marking clause boundaries, like strong variations of pitch at the end of constituents, pauses, lengthening the final syllable of words at the end of clauses, and part or whole repetitions of sentences (Cruttenden, 1994). These prosodic features facilitate segmentation of sentences according to syntactic structure and may highlight their structural characteristics. This kind of prosodic segmentation in natural language is in some ways similar to the manipulations in the present experiments: the prosodic features present in child-directed speech serve to focus the listener on the basic structural characteristics of grammar first, which once learned, allow the child to generalize to more complex structures. Indeed, the transition from one block of learning items with n levels of embedding to the next block with n+1 levels of embedding has a similar effect as these prosodic features in natural language, to stress the boundaries of embedded clauses.

Although most studies on child directed speech also mention short length as a feature of early sentences (Pine, 1994), some complexity-reducing features of child directed speech contribute to longer sentences rather than shortening them, such as repetition of constituents and lengthening the last syllable of a clause. This finding indicates the possibility that limited sequence length is a redundant feature of child directed speech playing a subordinate role in comparison to complexity-reducing features. Our data showed that the starting small ordering according to length failed to facilitate learning the sequences with embeddings, and possibly misdirect the learners' attention to string length as a relevant aspect of the grammar. These experimental and natural language studies together lend support to our proposal that the reduced computational complexity of the starting small regimen mainly is exploited by the young learner; whereas limited sentence length in child directed speech is a redundant feature that might even hamper the learning process when it is made salient at the cost of structural features, by the way the input is organized. This is an area that warrants further research.

Besides starting small, a number of recent studies with artificial languages suggest that certain extra-linguistic cues, also present in the natural situation, substantially ease the complexity of the learning task (Christiansen & Dale, 2001; Perruchet & Rey, 2005; Poletiek, 2002, 2006; Poletiek & Chater, 2006). First, the frequency distribution of a learning set may emphasize the structural properties of the underlying grammar. Poletiek and Chater (2006) and Poletiek (2006) showed that presenting more exemplars of less complex constructions had a positive effect on learning a finite state grammar, than presenting all types of constructions equally often. Perfors et al. (2011) gave their computational model an input with less frequent items having more complex structure than items having less complex structure. Similar to starting small, a frequency distribution following complexity may suggest a structure in the exemplars that mirrors the logical structure of the underlying grammar. Second, a primacy effect in AGL combined with a starting small input, also contributes to the explanation of the starting small cue. If learners (even adult learners) are better learners at the earlier stages of learning - as suggested by the primacy effect (Newport, Weiss, & Aslin, 2006) - starting small input makes it possible to acquire the basic pattern of a recursive rule in this sensitive stage. Lai and Poletiek's (2011) study showed the crucial importance of what is learned in the earliest stage, for eventually mastering a center-embedded structure.

Finally, other types of cues seem to play a role in grammar learning, and may possibly interact with starting small. For example, presentation modality influences performance in AGL tasks (Conway & Christiansen, 2005; Conway, Ellefson, & Christiansen, 2003; Saffran, 2002). Under some conditions, humans are better at encoding and processing auditory compared to visual input. In the present experiments, the stimuli were presented visually. Though auditory presentation, which best simulates most natural language learning situations, seems to be advantageous, the visual presentation format commonly used in AGL experimentation may possibly have enhanced the starting small effect for the self-embedded structure in Experiment 2 and 3, because each full exemplar could be viewed at once (Conway et al., 2003). In particular, long distance dependencies may be easier to recognize when the full string is visualized, as compared to the auditory presentation in which correct parsing of such a string depends on the memorization of previously heard elements for judging future ones. Future work must attempt to find out whether under more natural auditory conditions, recursive grammar learning is still differentially affected by starting small via staged input.

Interestingly, the role for these characteristics of the stimulus set, including the starting small effect demonstrated here, as a means for learning about complex structure may provide new insights into how recursive linguistic structure may be accommodated within stimulus based learning accounts. Our experimental data suggest an advantage for starting small when learning a grammar that incorporated recursive structure, and they show that this facilitation provided by the environment operates on the computational aspect of what has to be learned. These results are especially interesting because recursion is an important feature of natural language, and possibly of human cognition more broadly. Hence, ordering the input in a particular way may be crucial for learning to occur with complex patterns such as language. The current set of results have laid the groundwork for future experiments to explore the extent to which starting small may contribute to spoken language acquisition and inductive learning more generally.

Appendix A

Experiment 1: Learning exemplars of an artificial right branching recursive grammar (Figure 1a).

0 levels of embedding	1 level of embedding	2 levels of embedding
CW	CWPT	CKMWPH
CK	QZMW	QWXTMK
QZ	MKXH	MZSHCW
MW	PHQK	PVQZST
MK	XTSV	XHQKCZ
PH	SVCZ	STMWXV
PV	QKPT	CWXHSV
XT	QWCK	XVCKPT
SH	XTMZ	SHPTQZ
SV	CKQW	PTSVQW
	MZPV	MZPVXH
	XVPH	QKPHMZ
	PTCK	XTCZMK
	SHMW	PVCZSH
	QZST	QMMKXV
	STXH	MZQWCK
	CWXV	PHXTQK
	MKCZ	QZXVPT
	PVXT	STMWQZ
	SHQZ	MKSVCW

Experiment 1: Test materials

Grammatical	Ungrammatical
CZ	CT
QW	QH
QK	QV
MZ	MT
PT	PZ
XH	XW
XV	XK
ST	SZ
PHCW	QHCW
MWQK	MWXK
XHPT	XWPT
CZXV	CZXW
SVMZ	CVMZ
QWSH	QWKH
STPV	SZPV
MKCZ	MKCT
MZSVCW	XZSVCW
PVXHQK	PVMHQK
XTCZSH	XTCZQH
QWMKPT	QWMKPZ
CKMWXV	CKMTXV
PHSTMK	PZSTMK
SVQZXT	CVQZXT
STPHQZ	STCHQZ
XHPVCK	XHPVSK

Appendix B

Experiment 2: Learning exemplars of an artificial center embedding recursive grammar (Figure 1b).

0 levels of embedding	1 level of embedding	2 levels of embedding
CW	CPTW	CMPHWK
CK	QMWZ	QXMKTW
QZ	MXHK	MSCWHZ
MW	PQKH	PQSTZV
MK	XSVT	XQCZKH
PH	SCZV	SMXVWT
PV	QPTK	CXSVHW
XT	QCKW	XCPTKV
SH	XMZT	SPQZTH
SV	CQWK	PSQWVT
	MPVZ	MPXHVZ
	XPHV	QPMZHK
	PCKT	XCMKZT
	SMWH	PCSHZV
	QSTZ	QMXVKM
	SXHT	MQCKWZ
	CXVW	PXQKTH
	MCZK	QSPTVZ
	PXTV	SMQZWT
	SQZH	MSCWVK

Experiment 2: Test materials

Grammatical	Ungrammatical
CZ	CT
QW	QH
QK	QV
MZ	MT
PT	PZ
XH	XW
XV	XK
ST	SZ
PCWH	QCWH
MQKW	MXKW
XPTH	XPTW
CXVZ	CXWZ
SMZV	CMZV
QSHW	QKHW
SPVT	SPVZ
MCZK	MCTK
MSCWVZ	XSCWVZ
PXQKHV	PMQKHV
XCSHZT	XCQHZT
QMPTKW	QMPZKW
CMXVWK	CMXVTK
PSMKTH	PSMKTZ
SQXTZV	CQXTZV
SPQZHT	SCQZHT
XPCKVH	XPSKVH

Appendix C

Experiment 3: Learning and testing stimuli in all conditions: Starting Small according to number of levels of embedding, Starting Small according to length, and Randomly ordered. Squared brackets (not presented to participants) indicate embeddings.

Ordering according to Levels of Embedding			Ordering according to length			Random ordering		
	LoE	Length		Length	LoE		LoE	Length
QPK	0	3	CK	2	0	C[QP[CK]K]K	2	7
CWZ	0	3	SV	2	0	C[S[CWZ]V]WZ	2	8
QPWZ	0	4	CWZ	3	0	CWZ	0	3
CK	0	2	QPK	3	0	S[S[CK]V]V	2	6
SV	0	2	S[SV]V	4	1	S[QPWZ]V	1	6
C[CK]WZ	1	5	QPWZ	4	0	C[C[CK]WZ]K	2	7
C[CWZ]WZ	1	6	C[SV]WZ	5	1	QPK	0	3
QP[SV]WZ	1	6	C[CK]WZ	5	1	QP[SV]WZ	1	6
C[QPK]K	1	5	S[CWZ]V	5	1	C[C[CWZ]WZ]WZ	2	9
S[QPWZ]V	1	6	C[QPK]K	5	1	C[QPWZ]K	1	6
QP[QPK]K	1	6	QP[CWZ]K	6	1	C[QP[CWZ]K]WZ	2	9
S[CWZ]V	1	5	QP[QPK]K	6	1	S[S[SV]V]V	2	6
C[QPWZ]K	1	6	QP[SV]WZ	6	1	S[SV]V	1	4
QP[CK]WZ	1	6	S[C[SV]K]V	6	2	QP[CWZ]K	1	5
S[SV]V	1	4	QP[CK]WZ	6	1	QP[QP[QPWZ]WZ]WZ	2	12
C[SV]WZ	1	5	C[QPWZ]K	6	1	S[C[SV]K]V	2	6
QP[QPWZ]WZ	1	8	S[QPWZ]V	6	1	C[CK]WZ	1	5
QP[CWZ]K	1	6	C[C[SV]K]K	6	2	QP[QP[SV]WZ]K	2	9

QP[QPK]WZ	1	7	S[S[SV]V]V	6	2	S[QP[SV]WZ]V	2	8
C[QPWZ]WZ	1	7	S[S[CK]V]V	6	2	QP[C[CK]WZ]WZ	2	9
S[C[SV]K]V	2	6	C[CWZ]WZ	6	1	S[QP[CWZ]K]V	2	8
C[QP[QPK]WZ]WZ	2	10	C[S[SV]V]WZ	7	2	QP[QP[CWZ]K]K	2	9
S[QP[CWZ]K]V	2	8	QP[QPK]WZ	7	1	QP[CK]WZ	1	6
QP[C[SV]K]WZ	2	8	C[C[CK]WZ]K	7	2	S[C[CK]WZ]V	2	7
S[S[CK]V]V	2	6	C[S[QPK]V]K	7	2	C[CWZ]WZ	1	6
S[QP[SV]WZ]V	2	8	C[QP[CK]K]K	7	2	QPWZ	0	4
QP[S[CK]V]WZ	2	8	S[C[CK]WZ]V	7	2	C[QPWZ]WZ	1	7
S[C[CK]WZ]V	2	7	C[QPWZ]WZ	7	1	QP[S[QPWZ]V]K	2	9
QP[C[CK]WZ]WZ	2	9	S[C[QPWZ]K]V	8	2	C[SV]WZ	1	5
C[C[QPWZ]K]K	2	8	QP[S[CK]V]WZ	8	2	QP[C[SV]K]WZ	2	8
C[QP[SV]WZ]K	2	8	C[S[CWZ]V]WZ	8	2	S[CWZ]V	1	5
C[C[SV]K]K	2	6	QP[S[SV]V]WZ	8	2	QP[QPWZ]WZ	1	8
QP[S[QPWZ]V]K	2	9	QP[QPWZ]WZ	8	1	C[QP[QPK]WZ]WZ	2	10
QP[C[QPK]WZ]K	2	9	S[QP[CWZ]K]V	8	2	S[C[QPWZ]K]V	2	8
C[S[CWZ]V]WZ	2	8	QP[C[SV]K]WZ	8	2	QP[S[SV]V]WZ	2	8
QP[QP[CWZ]K]K	2	9	S[QP[SV]WZ]V	8	2	SV	0	2
S[QP[QPK]WZ]V	2	9	C[C[QPWZ]K]K	8	2	C[QP[SV]WZ]K	2	8
C[S[SV]V]WZ	2	7	C[QP[SV]WZ]K	8	2	QP[S[CK]V]WZ	2	8
S[C[QPWZ]K]V	2	8	S[S[QPWZ]V]V	8	2	CK	0	2
QP[QP[QPWZ]WZ]WZ	2	12	C[C[CWZ]WZ]WZ	9	2	C[C[QPWZ]WZ]WZ	2	10
C[C[CWZ]WZ]WZ	2	9	QP[QP[CWZ]K]K	9	2	C[QPK]K	1	5
C[QP[CWZ]K]WZ	2	9	QP[C[QPK]WZ]K	9	2	C[S[QPK]V]K	2	7
QP[S[SV]V]WZ	2	8	QP[S[QPWZ]V]K	9	2	QP[QPK]K	1	5
C[S[QPK]V]K	2	7	QP[QP[SV]WZ]K	9	2	C[C[QPWZ]K]K	2	8
C[C[QPWZ]WZ]WZ	2	10	C[QP[CWZ]K]WZ	9	2	QP[QPK]WZ	1	7
S[S[QPWZ]V]V	2	8	QP[C[CK]WZ]WZ	9	2	S[S[QPWZ]V]V	2	8

S[S[SV]V]V	2	6	S[QP[QPK]WZ]V	9	2	QP[C[QPK]WZ]K	2	9
QP[QP[SV]WZ]K	2	9	C[C[QPWZ]WZ]WZ	10	2	S[QP[QPK]WZ]V	2	9
C[QP[CK]K]K	2	7	C[QP[QPK]WZ]WZ	10	2	C[C[SV]K]K	2	6
C[C[CK]WZ]K	2	7	QP[QP[QPWZ]WZ]WZ	12	2	C[S[SV]V]WZ	2	7

Experiment 3: Test items with ungrammatical elements printed bold. Squared brackets were not presented to participants

Grammatical	Ungrammatical
SV	C**V**
CWZ	S**K**
QPK	QP**V**
QPWZ	S**WZ**
C[CK]K	C[C**V**]K
C[CWZ]K	C[SV]**V**
C[QPK]WZ	S[CK]**K**
C[SV]K	C[CWZ]**V**
QP[CK]K	QP[C**V**]K
QP[CWZ]WZ	QP[S**K**]K
QP[QPWZ]K	S[QP**V**]V
QP[SV]K	C[QP**V**]WZ
S[CK]V	QP[CWZ]**V**
S[QPK]V	QP[QP**V**]K
C[C[CWZ]K]K	C[C[CWZ]K]**V**
C[C[QPK]K]K	C[C[QPK]**V**]K
C[S[CK]V]WZ	C[S[CK]**K**]WZ
C[S[SV]V]K	C[S[S**WZ**]V]K
QP[C[CWZ]WZ]WZ	QP[C[SV]**V**]K
QP[C[SV]K]K	QP[S[SV]V]**V**
QP[QP[QPWZ]WZ]K	S[C[QPK]K]**K**
QP[QP[SV]WZ]WZ	S[QP[S**WZ**]K]V
QP[S[SV]V]K	QP[C[CWZ]WZ]**V**
S[C[QPK]K]V	QP[QP[SV]**V**]WZ
S[QP[SV]K]V	QP[QP[QPWZ]WZ]**V**

Chapter 6

Summary

Recursion is a crucial characteristic of the grammar of human languages (Chomsky, 1957; Corballis, 2007; Poletiek, 2011). Recently, the ability to process center-embedded recursion has been proposed to be the unique factor, distinguishing human from nonhuman beings (Hauser, Chomsky, & Fitch, 2002). Center-embedded structures, such as the following example: *the student that the teacher instructed improved*, are known for being difficult to understand and learn, since they produce long-distance hierarchical dependencies (e.g. *the student* […] *improved*). The present dissertation consists of four empirical studies, aiming to investigate the mechanism of learning and processing center-embedded recursive structures.

Chapter 2 addresses the question about the learnability of hierarchical center-embedded structures in two artificial grammar learning (AGL) experiments. In the AGL procedure, participants are first exposed to exemplars of an artificial grammar. Next, they give grammaticality judgments for new sequences that are either grammatical or ungrammatical. Participants' performance on this test task is an indication of how much they learned of the grammar from exposure to the exemplars. Experiment 1 showed that our participants could only learn the artificial language with a center-embedded rule, when they were exposed to a training input arranged according to increasing complexity. By contrast, participants, who received a randomly arranged training input, did not show any learning. Hence, there was a facilitating effect of "starting small" (SS). In the increased complexity condition (i.e. the SS condition), basic sentences *without* any embedding, i.e. zero level of embedding (0-LoE), were presented first, then one level of embedding (1-LoE) structures, and finally 2-LoE sentences.

In Experiment 2, we removed all 0-LoE learning items from the training phase. Therefore, participants were only trained with 1-LoE and 2-LoE items. We observed chance level performance, even with the SS ordering. Thus, the facilitation of SS disappeared when there was no sufficient exposure to the basic 0-LoE learning items. The results of Chapter 2 reveal that early and sufficient exposure to the basic simple structures (structures without embeddings, on which the recursive operation of inserting embedded clauses, can be applied) is a necessary condition for successful learning of complex center-embedded recursion. For natural grammar learning, this suggests that in order to learn how to parse and understand the sentence *the dog the man walks barks* properly, learners must have sufficient previous experience with the sentence *the dog barks.*

Chapter 3 further investigates characteristics of the training input that may facilitate learning of center-embedded structures. We tested a more sophisticated type of SS, which grows gradually (showing gradually higher LoE sentences in the input over time) rather than discretely. Moreover, we manipulated the frequency distribution of the input (equal versus unequal). The results of two experiments not only replicated the discretely clustered SS effect found in Chapter 2, but also showed a facilitative effect of the unequal frequency distribution. We found that the incremental SS ordering could enhance learning only when the frequency distribution of the training sentences was biased towards a higher frequency of 0-LoE training items, as compared to items with 1-LoE or 2-LoE. In addition, only participants, who received at least three consecutive blocks of 0-LoE training items during the earliest stage of exposure, performed above chance level. By contrast, participants, who were presented with a set containing both basic items without embeddings and a few items with embedded clauses at the beginning of exposure, did not show any learning. Having to deal right away with embedded sentences apparently disrupted the process of learning the embedding principle eventually. We conclude that the gradual SS input, together with a skewed frequency distribution, forming a combination that most resembles natural language acquisition (Kurumada, Meylan, & Frank, 2011; Poletiek & Chater, 2006), is optimal for learning. Our findings thus elaborate on the results of Chapter 2 and strengthen the view that early massive exposure to basic structures *without* any embedding is crucial for learning this complex syntactic pattern.

In Chapter 4 we investigated processing center-embedded structures in natural language. We examined the comprehension of recursive structures in natural language sentences (i.e. Dutch). In contrast to the standard locality view on processing embedded clauses (Gibson & Thomas, 1996; Kimball, 1973; Lewis, 1996), explaining processing difficulties by the distant positions of syntactically related words, we found that the congruency between the syntactic and semantic relatedness between the words had a much stronger effect on comprehension than their pure positions. Hierarchical center embedded structures were hardly more difficult to parse than linear right branching structures in our experiment. The mismatch between semantic and syntactic relations between the two words making a clause, however, strongly impeded comprehension. Thus, the long distance dependencies in the sentence *the dog the man walks barks* were easier to process by our participants than the short distance ones in the sentence *the dog walks the man who barks*. Our results offer a new perspective on the relative difficulty of processing hierarchical center-embedding, as compared to linear right-branching recursion. Our data suggest a new balance of the relative contributions to complex sentence processing of syntactic structure (hierarchical versus linear) on the one hand, and semantic content on the other hand, in favor of semantic influences.

In Chapter 5 we implemented an artificial linguistic stimulus environment to explore further under which circumstances SS would enhance learning. Experiment 1 showed that participants were able to learn right-branching recursive structures only when presented with a SS input (i.e. first 0-LoE, then 1-LoE, and finally 2-LoE). Experiment 2 replicated this SS effect with the more complex center-embedded type of recursion. However, since item complexity (number of LoE's) correlated perfectly with item length in these experiments, this confounder makes it difficult to conclude whether the SS effect was caused by the increasing LoE's or increasing item length. Experiment 3 addressed this question by disentangling the two factors complexity and length of sentence. We found that participants were able to learn the center-embedded structures when they were exposed to staged input with increasing LoE's, but not when they were exposed to a training regimen of the same sentences increasing in length. In conclusion, our three experiments showed what we also showed in Chapter 2 and 3, that a SS regimen facilitates learning both the right-branching and the center-embedded recursive grammar. More specifically, the data of Chapter 5 suggest that it is the organization of the input in terms of increasing item complexity (not increasing item length) that is effective in this facilitation.

In sum, the results of the present dissertation may help us understand the cognitive mechanism of processing and learning center-embedded recursion: an issue that is much debated in the study of language. Globally, the results of the present research generated two novel hypotheses about this process. First, learning the self-referential structure of recursion is highly conditional upon the organization over time of the input, and in particular, early and intensive exposure to basic sentences *without* any recursive loop (i.e. simple sentences without relative clauses and adjacent dependencies). Second, our data stress the crucial importance of the *semantic* associativeness between the elements that are syntactically related but positioned far away from each other in the sentence. It is not the positional distance between elements, per se, but their semantic "distance" that mainly determines how easily they are unified in the parsing process. Overall, our research suggests that *the fit* between the linguistic environment (SS), semantic memory and structural complexity (center-embedded recursion) determines learning and processing recursion in language, rather than these factors individually.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*(8), 1067-1074.

Abney, S. P., & Johnson, M. (1991). Memory Requirements and Local Ambiguities of Parsing Strategies. *Journal of Psycholinguistic Research, 20*(3), 233-250.

Altmann, G. T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15250-15251.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Arciuli, J., & Torkildsen, J. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in psychology, 3*, 324.

Ayoub, C., Vallotton, C. D., & Mastergeorge, A. M. (2011). Developmental Pathways to Integrated Social Skills: The Roles of Parenting and Early Intervention. *Child Development, 82*(2), 583-600.

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

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes, 1*, 249-262.

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

Bahlmann, J., & Friederici, A. D. (2006). FMRI investigation of the processing of simple linear and embedded hierarchical structures: An artificial grammar task. *Journal of Cognitive Neuroscience*, Annual Meeting Supplement, p.126.

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2006). Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience, 18*(11), 1829-1842.

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage, 42*(2), 525-534.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.

Baum, S. R. (1993). Processing of Center-Embedded and Right-Branching Relative Clause Sentences by Normal Elderly Individuals. *Applied Psycholinguistics, 14*(1), 75-88.

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport, 23*(3), 139-145.

Bellinger, D. (1980). Consistency in the pattern of change in mothers' speech: Some discriminant analyses. *Journal of Child Language, 7*, 469-487.

Berwick, R. C., Beckers, G. J., Okanoya, K., & Bolhuis, J. J. (2012). A Bird's Eye View of Human Language Evolution. *Frontiers in evolutionary neuroscience, 4*, 5.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences, 15*(3), 113-121.

Blauberg, M. S., & Braine, M. D. S. (1974). Short-Term-Memory Limitations on Decoding Self-Embedded Sentences. *Journal of Experimental Psychology, 102*(4), 745-748.

Bloomfield, T. C., Gentner, T. Q., & Margoliash, D. (2011). What birds have to say about language. *Nature Neuroscience, 14*(8), 947-948.

Blumenthal, A. L., & Boakes, R. (1967). Prompted Recall of Sentences. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 674-676.

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

Brysbaert, M. (2007). The Language-As-Fixed-Effect-Fallacy: Some Simple SPSS Solutions to a Complex Problem. London: Royal Holloway, University of London.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*. Cambridge, MA:MIT Press.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*(6), 500-508.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chater, N., & Vitányi, P. (2002). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*, 19-22.

Chater, N., & Vitanyi, P. (2007). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology, 51*(3), 135-163.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

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

Chomsky, N. (1980). Rules and Representations. *Behavioral and Brain Sciences, 3*(1), 1-15.

Chomsky, N. (1995). *The minimalist program*. MIT, Cambridge.

Chomsky, N., & Miller, G. (1963). Introduction to the Formal Analysis of Natural Languages. *Handbook of mathematical Psychology,* R. Luce et al., eds. New York: John Wiley, 286-287.

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

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(5), 489-509.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes, 27*, 231-256.

Christiansen, M. H., & Dale, R. A. C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In J.D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M. H. & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning, 59 (Suppl. 1*), 126-161.

Church, K. (1982). *On memory limitations in natural language processing*. Bloomington, IN: Indiana University Linguistics Club.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804-809.

Cochran, B. P., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal* *of Memory and Language, 41*(1), 30-58.

Cohen, L. B., & Strauss, M. S. (1979). Concept-Acquisition in the Human Infant. *Child Development, 50*(2), 419-424.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24-39.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When less is less and when less is more: Starting small with staged input. In R. Alterman & D. Kirsh (Eds.), In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 270-275). Mahwah, NJ: Erlbaum.

Coolidge, F. L., Overmann, K. A., & Wynn, T. (2011). Recursion: what is it, who has it, and how did it evolve? *Wiley Interdisciplinary Reviews-Cognitive Science, 2*(5), 547-554.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science, 31*(4), 697-704.

Craik, F. I. M., & Tulving, E. (1975). Depth of Processing and Retention of Words in Episodic Memory. *Journal of Experimental Psychology-General, 104*(3), 268-294.

Cruttenden, A. (1994). Phonetic and prosodic aspects of baby talk. In C. Gallaway and B. Richards (eds), Input and Interaction in Language Acquisition. Cambridge, UK: Cambridge University Press, 135-152.

de Vries, M. H., Christiansen, M. H. & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics, 5*, 10-35.

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition, 107*(2), 763-774.

de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: evidence from sequence learning. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1598), 2065-2076.

Eenshuistra, R. M., Weidema, M. A., & Hommel, B. (2004). Development of the acquisition and control of action-effect associations. *Acta Psychologica, 115*(2-3), 185-209.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L. (1991). Incremental learning, or the importance of starting small. *Technical Report, 9101*. Center for Research in Language, University of California at San Diego.

Elman, J. L. (1993). Learning and Development in Neural Networks - the Importance of Starting Small. *Cognition, 48*(1), 71-99.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research, 52*(2), 321-335.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Piraha - Another look at the design features of human language. *Current Anthropology, 46*(4), 621-646.

Fedor, A., Varga, M., & Szathmary, E. (2012). Semantics boosts syntax in artificial grammar learning tasks with recursion. *Journal of experimental psychology. Learning, memory, and cognition, 38(3)*, 776-782.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15.

Fitch, W. T. (2011). Unity and diversity in human language. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1563), 376-388.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*(5656), 377-380.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179-210.

Fletcher, J., Maybery, M. T., & Bennett, S. (2000). Implicit learning differences: A question of development level? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 246-252.

Fodor, J. A., & Garrett, M. (1967). Some Syntactic Determinants of Sentential Complexity. *Perception & Psychophysics, 2*(7), 289-296.

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *Neuroimage, 32*(2), 956-967.

Foss, D. J., & Cairns, H. S. (1970). Some Effects of Memory Limitation Upon Sentence Comprehension and Recall. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 541-547.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences, 8*(6), 245-247.

Friederici, A. D. (2012). Language development and the ontogeny of the dorsal pathway. *Frontiers in evolutionary neuroscience, 4*, 3.

Friederici, A. D., Bahlmann, J., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion. *Biolinguistics, 5*(1-2), 87-104.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2458-2463.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C.E. Snow & C.A. Ferguson. *Talking to children: Language input and acquisition*, 63-68. Cambridge, England: Cambridge University Press.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language, 58*, 161–187.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204-1207.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Gibson, E., & Thomas, J. (1996). *The processing complexity of English center-embedded and self-embedded structures.* Paper presented at the the NELS, University of Massachusetts, Ahmerst:Graduate Student Association.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*(3), 225-248.

Gold, E. (1967). Language identification in the limit. *Information and Control, 10* (5), 447–474.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences, 7*(5), 219-224.

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

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

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

Gomez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. Infancy, 7(2), 183-206.

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic Processing in Baboons (Papio papio). *Science, 336*(6078), 245-248.

Hakuta, K. (1981). Grammatical Description Versus Configurational Arrangement in Language-Acquisition - the Case of Relative Clauses in Japanese. *Cognition, 9*(3), 197-236.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science, 14*, 31–38.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Helmers, C., & Patnam, M. (2011). The formation and evolution of childhood skill acquisition: Evidence from India. *Journal of Development Economics, 95*(2), 252-266.

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 213-231). Chichester: Wiley.

Hildebrandt, N., Caplan, D., & Evans, K. (1987). The Mani Left Ti without a Trace - a Case-Study of Aphasic Processing of Empty Categories. *Cognitive Neuropsychology, 4*(3), 257-302.

Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science, 32*, 1021-1036.

Holmes, V. M. (1973). Order of Main and Subordinate Clauses in Sentence Perception. *Journal of Verbal Learning and Verbal Behavior, 12*(3), 285-293.

Hudson, R. (1996). The difficulty of (so-called) self-embedded structures. *UCL Working Papers in Linguistics 8*, 283-314.

Johnson, M. (1998). Proof Nets and the Complexity of ProcessingCenter-Embedded Constructions. *Journal of Logic, Language and Information 7*(4), 433-447.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*(1), 61-112.

Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension - Individual-

Differences in Working Memory. *Psychological Review, 99*(1), 122-149.

Karbach, J., Kray, J., & Hommel, B. (2011). Action-effect learning in early childhood: does language matter? *Psychological Research-Psychologische Forschung, 75*(4), 334-340.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology-General, 126*(3), 278-287.

Karlsson, F. (2007). Constraints on multiple center embedding of clauses. *Journal of Linguistics, 43,* 365–392.

Karlsson, F. (2010). Syntactic recursion and iteration. In H. v. d. Hulst (Ed.), *Recursion and Human Language* (pp. 43–68). Berlin/New York: Mouton de Gruyter.

Kersten, A. W., & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language, 44*(2), 250-273.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes, 22*(6), 860-897.

Kidd, E., Lieven, E., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences, 32*, 132-142.

Kimball, J. (1973). 7 Principles of Surface Structure Parsing in Natural Language. *Cognition, 2*(1), 15-47.

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology-General, 132*(4), 551-565.

Knowlton, B. J., & Squire, L. R. (1994). The Information Acquired during Artificial Grammar Learning. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(1), 79-91.

Krueger, K. A., & Dayan, P. (2009). Flexible shaping: How learning in small steps helps. *Cognition, 110*, 380-394.

Kurumada, C., Meylan, S., & Frank, M. C. (2011). Zipfian word frequencies support statistical word segmentation. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp.2667-2672). Austin, TX: Cognitive Science Society.

Lai, J., & Poletiek, F. H. (2010). *The impact of starting small on the learnability of recursion.* Paper presented at the Proceddings of the 32rd Annual Conference of the Cognitive Science Society, Austin.

Lai, J. & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118(2),* 265-273.

Lany, J., Gomez, R. L., & Gerken, L. (2007) The role of prior experience in language acquisition. *Cognitive Science, 31*, 481-507.

Larkin, W., & Burns, D. (1977). Sentence Comprehension and Memory for Embedded Structure. *Memory & Cognition, 5*(1), 17-22.

Lewis, R. L. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Review, 25,* 93-116.

Lobina, D. J. (2011). Recursion and the competence/performance distinction in AGL tasks. *Language and Cognitive Processes, 26*(10), 1563-1586.

Locker, L., Hoffman, L., Bovaird, J. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39*(4), 723-730.

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypotheses. In L.R. Gleitman & A.K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 812-817), Mahwah, NJ: Lawrence Erlbaum.

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.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review, 101*(4), 676-703.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 249–308). Hillsdale, NJ:Erlbaum.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*(1), 50-68.

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language, 54*(4), 466-490.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the United States of America, 106*(20), 8362-8367.

Mandler, J. M., & Mcdonogh, L. (1993). Concept-Formation in Infancy. *Cognitive Development, 8*(3), 291-318.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77-80.

Marcus, M. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

Marks, L. E. (1968). Scaling of Grammaticalness of Self-Embedded English Sentences. *Journal of Verbal Learning and Verbal Behavior, 7*(5), 965-967.

McDonald, J. L., & Plauche, M. (1995). Single and correlated cues in an artificial language learning paradigm. *Language and Speech, 38 (3)*, 223-236.

Miller, G. A. (1962). Some Psychological Studies of Grammar. *American Psychologist, 17*(11), 748-762.

Miller, G. A., & Isard, S. (1964). Free Recall of Self-Embedded English Sentences. *Information and Control, 7*(3), 292-303.

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

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2009). Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. In N. A. Taatgen & H. v. Rijn (Eds.), *Proceedings of the 31st* *annual conference of the Cognitive Science Society* (pp. 177–182). Austin, TX: Cognitive Science Society.

Morrill, G. (2000). Incremental processing and acceptability. *Computational Linguistics, 26*(3), 319-338.

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of Embedded Syntactic Structures Depends on Prosodic Cues. *Cognitive Science, 34*(2), 338-349.

Nelson, K. (1985). *Making Sense: The Acquisition of Shared Meaning*. Orlando, FL: Academic Press.

Newell, K. M. (1991). Motor Skill Acquisition. *Annual Review of Psychology, 42*, 213-237.

Newmeyer, F. (1988). Extensions and implications of linguistic theory: an overview. In F. Newmeyer, (ed.) *Linguistics: The Cambridge Survey 2. Linguistic Theory: Extensions and Implications*. Cambridge: Cambridge University Press. 1-14.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*, 147-172.

Newport, E. L. (1990). Maturational Constraints on Language-Learning. *Cognitive Science, 14*(1), 11-28.

Newport, E.L., Weiss, D.J., & Aslin, R.N. (2006). Statistical learning in language, vision and comparative cognition. What have we learned, where are we going? In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (p. 2662). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1047-1052). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Perfors, A., & Burns, N. (2010). Adult language learners under cognitive load do not over-regularize like children. In S. Ohlsson & R. Catrambone (eds.) *Proceeding of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2524-2529). Austin, TX: Cognitive Science Society.

Perfors, A., Tenenbaum, J.B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118(3)*, 306-338.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review, 12*(2), 307-313.

Pickering, M., & Barry, G. (1991). Sentence Processing without Empty Categories. *Language and Cognitive Processes, 6*(3), 229-259.

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & B.J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 109-149). Cambridge, UK: Cambridge University Press.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure. Learning, development, and conceptual change*. Cambridge, MA: MIT Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*(2), 201-236.

Plunkett, K., & Marchman, V. (1990). *From rote learning to system building*. Center for Research in Language. TR 9020, University of California, San Diego.

Plunkett, K., & Marchman, V. (1993). From Rote Learning to System Building - Acquiring Verb Morphology in Children and Connectionist Nets. *Cognition, 48*(1), 21-69.

Poletiek, F. H. (2002). Implicit learning of a recursive rule in an artificial grammar. *Acta Psychologica, 111*(3), 323-335.

Poletiek, F. H. (2006). Representative sampling in an artificial grammar learning task. In P. Juslin, & K. Fiedler (Eds.), *Information sampling as adaptive cognition* (pp. 440-455). Cambridge, UK: Cambridge University Press.

Poletiek, F. H. (2011). What in the World Makes Recursion so Easy to Learn? A Statistical Account of the Staged Input Effect on Learning a Center-Embedded Structure in Artificial Grammar Learning (AGL). *Biolinguistics, 5*(1-2), 36-42.

Poletiek, F. H., & Chater, N. (2006). Grammar induction proWts from representative stimulus sampling *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1968–1973). Mahwah, NJ: Lawrence Erlbaum Associates.

Poletiek, F.H., Chater, N., & Van den Bos, E., (Submitted). The influence of stimulus frequencies in Artificial Grammar Learning. Evidence for deep and surface statistical learning. *Manuscript submitted for publication*.

Poletiek, F. H., & Fitz, H. (submitted). What baboons might tell us about center-embedded structures in human language.

Poletiek, F. H., & Lai, J. (2012). How Semantic Biases in Simple Adjacencies affect Learning a Complex Structure with Non-Adjacencies in AGL. A Statistical Account. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1598), 2046-2054.

Poletiek, F.H., & Van Schijndel, T.J.P. (2009) Stimulus Set Size and Grammar Coverage in Artificial Grammar Learning. *Psychonomic Bulletin and Review 16(6*), 1058-1064.

Poletiek, F.H., & Wolters, G. (2009). What is learned about fragments in artificial grammar learning? A transitional probabilities approach. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62*, 868-876.

Powell, A., & Peters, R. G. (1973). Semantic Clues in Comprehension of Novel Sentences. *Psychological Reports, 32*(3), 1307-1310.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for Representations of Perceptually Similar Natural Categories by 3-Month-Old and 4-Month-Old Infants. *Perception, 22*(4), 463-475.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*(1), 1-23.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855-863.

Reber, A. S. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology-General, 118*(3), 219-235.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology-General, 125*(2), 123-138.

Reich, P. (1969). The finiteness of natural language. *Language, 45*, 831–843.

Reuland, E. (2010). Imagination, Planning, and Working Memory The Emergence of Language. *Current Anthropology, 51*, S99-S110.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (Papio Papio). *Cognition, 123*(1), 180-184.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition, 72*(1), 67-109.

Rohde, D. L. T., & Plaut, D. C. (2003). Less is less in language acquisition. In P. Quinlin (Ed.), *Connectionist modelilng of cognitive development* (pp. 189-231). Ove, UK: Psychology Press.

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive, hierarchical grammatical structures. *PLoS ONE 7*(10): e45885. doi:10.1371/journal.pone.0045885.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Schiller, N. O. (2008). The gift of tongues: A neurocognitive approach to articulatory-motor learning. In I. v. Keulen (Ed.), *Brain visions. How the brain sciences could change the way we eat, communicate, learn and judge* (pp. 248-255). The Hague: STT (STT 73).

Schiller, N. O., & Meyer, A. S. E. (2003). *Phonetics and phonology in language comprehension and production. Differences and similarities.* Berlin, New York: Mouton de Gruyter.

Schlesinger, I. M. (1975). Why a sentence in which a sentence in which a sentence is embedded is embedded is difficult. *Linguistics, 153*, 53–66.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. Clark (Eds.), *The development of movement control and coordination* (pp. 113-150). New York: Wiley.

Skinner, B. F. (1957). *Verbal behavior.* . New York: Appleton-Century-Crofts.

Slobin, D.I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behavior, 5(3), 219-227.

Snow, C. E. (1986). Conversations with children. In P. F. a. M. Garman (Ed.), *Language Acquisition*. Cambridge: Cambridge University Press.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Hillsdale, NJ: Lawrence Erlbaum.

Starkey, D. (1981). The Origins of Concept-Formation - Object Sorting and Object Preference in Early Infancy. *Child Development, 52*(2), 489-497.

Stolz, W. S. (1967). A Study of Ability to Decode Grammatically Novel Sentences. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 867-873.

Tallerman, M., Newmeyer, F., Bickerton, D., Bouchard, D., Kaan, E., & Rizzi, L. (2009). What kinds of syntactic phenomena must biologists, neurobiologists, and computer scientists try to explain and replicate? In D. Bickerton & E. Szathmáry (Eds.), *Biological Foundations and Origin of Syntax* (Vol. 3, pp. 135-160). Cambridge, MA: MIT Press.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition.

Cognitive Linguistics, 11(1-2), 61-82.

Tomasello, M. (2003). *Constructing a language: a usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Udden, J., Araujo, S., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2009). A matter of time: Implicit acquisition of recursive sequence structures. In N. Taatgen & H. V. Rijn (Eds.), *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (pp. 2444-2449). Austin, TX: Cognitive Science Society.

van den Bos, E., & Poletiek, F. H. (2010). Structural selection in implicit learning of artificial grammars. *Psychological research, 74*(2), 138-151.

van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative récursive syntactic structures by a songbird species. *Proceedings of the National Academy of Science, U.S.A., 106*, 20538-20543.

Vasishth, S. (2001). An empirical evaluation of sentence processing models: Center embeddings in Hindi. In M. Daniels, D. Dowty, A. Feldman, & V. Metcalf (Eds.), *OSUWPL, Volume 56*, 159-181, Ohio State University.

Vosse, T., & Kempen, G. (1991). *A hybrid model of human sentence processing: parsing rightbranching, center-embedded and cross-serial dependencies.* Paper presented at the The Second International Workshop on Parsing Technologies, Cancun, Mexico.

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior, 9,* 398–404.

Weckerly, J., & Elman, J. (1992). *A PDP approach to processing center-embedded sentences.* Paper presented at the The Fourteenth Annual Meeting of the Cognitive Science Society, Hillsdale, NJ: Lawrence Erlbaum Associates.

Wells, J., 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,* 250–271.

Samenvatting

(Summary in Dutch)

Recursie is een cruciaal kenmerk van de grammatica’s van natuurlijke talen (Chomsky, 1957; Corballis, 2007; Poletiek, 2011). Recentelijk is voorgesteld dat het kunnen verwerken van *center-embedded CE* (in het midden ingebedde) recursieve structuren (CE) een uniek menselijk vermogen is (Hauser, Chomsky, en Fitch, 2002). Sommige aapsoorten, die wel eenvoudige sequentiële structuren kunnen leren, zijn niet in staat om een CE regel te leren. De zin ‘de student, die de leraar instrueerde, verbeterde’, is een natuurlijk voorbeeld van deze vorm van recursie. Hoewel in principe leerbaar, is CE recursie ook in natuurlijke taal soms moeilijk te begrijpen; de meeste theorieën verklaren die moeilijkheid door de lange afstand –hiërarchische- afhankelijkheden in een CE-zin (zoals bijvoorbeeld tussen ‘de student’ en ‘verbeterde’). Dit proefschrift rapporteert vier empirische onderzoeken over het mechanisme van het leren en verwerken van CE recursieve structuren.

Hoofdstuk 1 gaat in op de voorwaarden waaronder hiërarchische CE structuren (beter) worden geleerd. In twee experimenten met het kunstmatige grammatica’s “Artificial Grammar Learning”(AGL)-paradigma, wordt het effect van eigenschappen van de leeromgeveing onderzocht. In de AGL-procedure, krijgen de deelnemers eerst een aantal voorbeeldzinnen van een kunstmatige grammatica te zien. Vervolgens geven de deelnemers grammaticaliteits-oordelen over nieuwe zinnen die ofwel grammaticaal correct of ongrammaticaal zijn. De accuraatheid van deze oordelen is een indicatie van hoeveel de deelnemers hebben geleerd over de grammatica, van de voorbeeldzinnen. In experiment 1 konden de deelnemers de kunstmatige taal met CE-recursie, alleen leren wanneer de leer-zinnen waren geordend naar toenemende complexiteit. De complexiteit nam toe met het aantal inbeddingen in de zin: eerst werden zinnen zonder inbedding (0-LoE: 0-Level of Embedding) gepresenteerd, vervolgens werd één niveau van inbedding (1-LoE) toegevoegd en tenslotte werd een tweede niveau van inbedding (2-LoE) toegevoegd. Deelnemers die werden blootgesteld aan een verzameling zinnen die in willekeurige volgorde stonden, konden de kunstmatige taal niet leren. Kortom, er was een leervoordeel van "klein beginnen"*“Starting Small”* (SS). In Experiment 2, hebben we alle 0-LoE structuren uit de trainingsfase verwijderd. De deelnemers kregen dus alleen 1-LoE en 2-LoE structuren te zien. Dit had tot gevolg dat het leervoordeel verdween, zelfs wanneer de SS volgorde werd geïmplementeerd. Dus het voordeel van ordening volgens het SS principe verdween wanneer de deelnemers onvoldoende werden blootgesteld aan de basis 0-LoE structuren. Het door ons gevonden hoge leereffect van 0-LoE zinnen is opmerkelijk, omdat in díe zinnen de recursieve structuur van inbedding juist niet tot uiting komt. Dit suggereert dat ook bij het leren van natuurlijke grammatica, voldoende basisstructuren moeten worden aangeboden zonder bijzinnen om de bijzin-constructie te leren. Dus, voor het ontleden en begrijpen van de zin ‘De hond die door de man uitgelaten wordt, blaft’ is het nodig dat er voldoende voorkennis is over de zin; ‘De hond blaft’. Deze hypothese hebben wij onderzocht in hoofdstuk 3.

In hoofdstuk 2 zijn de leerbevorderende eigenschappen van de input tijdens de trainingsfase verder onderzocht. Daarvoor hebben we een meer geavanceerde vorm van SS getest, namelijk één die geleidelijk toeneemt in plaats van discreet (geleidelijk worden steeds hogere LoE zinnen in de input ingevoegd). Bovendien hebben we de frequentieverdeling van de input (gelijk versus ongelijk) gemanipuleerd. Met de twee experimenten hebben we de resultaten van het SS effect uit hoofdstuk 1 kunnen repliceren. Daarnaast vonden we een faciliterend effect van de ongelijke frequentieverdeling. We vonden dat een geleidelijke ordening van SS het leerproces alleen positief beïnvloedde, wanneer de frequentieverdeling van de te leren zinnen scheef was in die zin dat zinnen met 0-LoE vaker voorkwamen dan complexere zinnen met 1-LoE, en die weer vaker dan zinnen met 2-LoE. Het bleek dat alleen die deelnemers, die ten minste drie opeenvolgende blokken van zinnen met 0-LoE gepresenteerd kregen tijdens het beginstadium van de leerfase, boven kansniveau presteerden. Deelnemers die van meet af aan een combinatie van zowel basiszinnen zonder inbeddingen, als ingebedde zinnen gepresenteerd kregen, leerden niets. Het direct blootstellen aan ingebedde zinnen verstoort blijkbaar het proces van het leren van het inbedding principe. We concluderen dat de geleidelijke SS-input, samen met een scheve frequentieverdeling, (een combinatie die het meest lijkt op de natuurlijke taalverwerving; Kurumada, Meylan, & Frank, 2011; Poletiek & Chater, 2006), optimaal is voor het leren. Onze bevindingen zijn een uitbreiding op de resultaten van hoofdstuk 1 en ondersteunen de hypothese dat vroege intensieve blootstelling aan basisstructuren *zonder* inbedding, van cruciaal belang is voor het leren van dit complexe syntactische patroon.

In Hoofdstuk 3 hebben we onderzoek gedaan naar het verwerken van ingebedde structuren in natuurlijke taal (Nederlands). In tegenstelling tot de standaard ‘locatie visie’ op de verwerking van ingebedde zinnen (Gibson & Thomas, 1996; Kimball, 1973; Lewis, 1996), die verwerkingsproblemen verklaart door de grote afstand tussen syntactisch gerelateerde woorden, vonden we dat de congruentie (*match*) tussen de syntactische en semantische gerelateerdheid tussen de woorden, een veel sterkere invloed op begrip had, dan de afstand tussen de woorden. In ons experiment waren hiërarchische CE structuren nauwelijks moeilijker te ontleden dan lineaire rechts-vertakkende structuren. De *mismatc*h tussen de semantische en syntactische relatie tussen woorden die deel uitmaakten van een bijzin, hinderden echter wel ernstig het begrip. Zo konden de deelnemers de lange-afstand in de zin: *“de hond die door de man uitgelaten wordt, blaft”* gemakkelijker verwerken, dan de korte-afstand in de zin: *“de hond laat de man uit die blaft”*. Deze resultaten bieden een nieuw perspectief op de relatieve moeilijkheid van het verwerken van hiërarchische CE, in vergelijking met lineaire rechts-vertakkende recursie. Onze gegevens suggereren een nieuw evenwicht van de relatieve bijdrage aan complexe zinsverwerking van syntactische structuur (hiërarchische versus lineair) enerzijds en semantische inhoud anderzijds, ten gunste van semantische invloeden.

In hoofdstuk 4 hebben we een kunstmatige taal geïmplementeerd om zodoende verder te het SS effect te verkennen. In lijn met de resultaten van Hoofdstuk 1, tonen de resultaten van Experiment 1 aan dat de deelnemers rechts vertakkende recursieve structuren alleen konden leren wanneer de input volgens het SS-principe werd gepresenteerd (dus eerst 0-LoE, dan 1-LoE, en ten slotte 2-LoE). In experiment 2 werd het SS effect gerepliceerd maar nu met de CE vorm van recursie. Aangezien echter in deze experimenten (en in de experimenten van Hoofdstuk 1) complexiteit (het aantal LoE's) volledig gecorreleerd is met zinslengte, is het moeilijk te concluderen of het SS-effect veroorzaakt werd door de toenemende LoE’s of door toename van zinslengte. In experiment 3 zijn de factoren complexiteit en zinslengte daarom losgekoppeld. Zinnen met twee LoE’s konden korter zijn dan zinnen met 1 LoE. We vonden dat de deelnemers de CE-structuren beter leerden wanneer ze werden blootgesteld aan een leerinput die groeide in complexiteit, maar niet beter leerden van een input die groeide volgens zinslengte. Samenvattend tonen de drie experimenten opnieuw aan dat SS het leren van recursieve grammatica vergemakkelijkt. Meer in het bijzonder laten de uitskomsten van hoofdstuk 4 zien dat dit ook geldt voor het leren van rechts-vertakkende structuren. De gegevens van hoofdstuk 4 suggereren bovendien dat SS alleen een faciliterend effect op het leren heeft wanneer het leerregime georganiseerd is volgens toenemende complexiteit van de zinnen, maar niet bij enkele toename van zinslengte.

De resultaten van dit proefschrift helpen het cognitieve mechanisme te begrijpen dat verantwoordelijk is voor het verwerken en leren van CE recursie: een veelbesproken en relatief onbegrepen aspect van taalverwerving. Globaal gezien hebben de resultaten van ons onderzoek twee volledig nieuwe hypothesen gegenereerd over dit proces. Ten eerste blijkt de leerbaarheid van recursie in de context van taalverwerving zeer afhankelijk te zijn van hoe de stimulus-input is georganiseerd. In het bijzonder is het van belang dat vroege en intensieve blootstelling aan basiszinnen zonder enige recursieve lus (in natuurlijke taal: eenvoudige zinnen zonder bijzinnen) plaatsvindt. Ten tweede benadrukken onze gegevens het cruciale belang van de semantische gerelateerdheid tussen elementen met een syntactisch verband, maar die ver van elkaar in de zin zijn geplaatst. Het is niet de positionele afstand tussen de elementen, per se, maar hun “semantische afstand" die lijkt te bepalen hoe gemakkelijk ze kunnen worden geïntegreerd in het ontledingsproces. Over het geheel genomen suggereert ons onderzoek dat het samenspel tussen de linguïstische omgeving (SS), het semantisch geheugen en de structurele complexiteit (CE- recursie) het leren en verwerken van recursie in taal bepaalt, en niet een van deze factoren alleen, noch elke factor afzonderlijk.

Acknowledgements

A four-year PhD project is like an Olympic game, but only for one player. In the long fascinating journey of competing with myself to achieve the ultimate goal, I could not have succeeded without the support of so many great people.

First, I would like to express my sincere appreciation to dr Fenna Poletiek, who has given me thorough insightful scientific guidance through this research. I am very grateful to Prof. Niels Schiller. Since being the first master student of Niels when he moved to Leiden, I have been inspired and highly motivated by him to explore the interesting field of psycholinguistics. I am also thankful to the members of my PhD committee, Prof. Padraic Monaghan, Prof. Pienie Zwitserlood, Prof. Carel ten Cate, dr Wido La Heij and dr Yiya Chen, for your interest in my research.

I benefit a lot from inspiration and insight from the LACG group by Niels and other linguistic colleagues. I appreciate insightful discussions during labmeetings from Bernhard and other cognitive psychology members. My time in the office has been made so enjoyable by friendly and joyful colleagues, Mischa, Henk, Xin, Lesya, Sharon, Mikael, Pascal, Stephan, Roy, Kerwin, Esther, Michiel and André. Especially, I want to thank my former officemates (Marieke and Rinus) and current ones (Anne and Kalinka) for all the hilarious, unforgettable and crazy moments that we share. I feel lucky to be surrounded by all your love and devotion. Special thanks also go to Stephen. He plays not only the role of my daily Dutch instructor, who kindly teaches me the beautiful language and culture in an amusing way, but also a great trustful friend, who shares all the ups and downs. Albertien, Atie and Marianne, thank you for taking care of me during the past four years. Many thanks for Jiska for your English and Dutch editing suggestions! I also appreciate a lot for the technical support for using Praat by colleagues from Linguistics, Jos and Daan. Dear colleagues from Scriptie Atelier, thanks for providing such a pleasant working atmosphere!

I want to thank all my amazing friends, though it is difficult to acknowledge everyone here. Especially, my best friends, Ao Chen and Xiaohui, thanks for always being there for me. Chenguang, thank you sincerely for your love and support during this long journey!

Finally, I would like to express my wholehearted gratitude to my parents, who both work in a university. They taught me not only knowledge, but also how to be a good person. The first English word that they taught me was “kite” when I was three years old. After growing up, I travelled to different corners of the world like a kite. No matter where I go, I miss them deeply. They raised me with love and caring, so I could see the beauty of the world on their shoulders.

致我敬爱的父母
赖爱胜、蔡琳

漫步人生路，博士四年犹如一场四年一届的奥林匹克运动会，然而不同的是，这是一场一个人的奥林匹克，是自我挑战的一场竞赛，是从优秀到卓越的精彩一课。因为有梦想，所以远航；因为有希冀，所以登顶。在攀登这座人生高峰的旅途中，是亲朋好友的关切和支持给我力量，更是父母无尽的爱让我每一个脚印都走得那么坚实、那么温暖。

在这篇博士论文结尾最重要的部分，我想向我最亲爱的父母致谢！我的父母都是高校教师，他们教给我的不仅是知识与获得知识的方法，更是通过自己的言传身教让我耳濡目染学会如何去爱，如何做一个优秀的人。在我三岁时父母教会我的第一个英文单词便是“kite” (风筝)，长大后我便如同一只风筝飞到世界不同角落。然而，纵使飞得再远，心中牵挂的永远是家，忘不了每次出发前妈妈一针一线地为我加固衣服，忘不了爸爸情深意切地叮咛嘱咐。爸爸妈妈用无尽的爱呵护、养育着我，给了我一个幸福完美的家。亲爱的爸爸妈妈，您们用大爱将我高高托起，于是我才能在您们坚实的肩膀上看到世界的美！我爱您们！仅以此篇博士论文献给您们！

Curriculum Vitae

Jun Lai was born on August 1st, 1982, in Ganzhou City, China. In 2000, she completed high school education in Ganzhou Nr. Three Middle School, and entered Sun Yat-sen University (Guangzhou, China) with a high scholarship for outstanding freshman, covering four years' tuition fee and living expenses. From 2000-2004, she won the annual scholarship consecutively. In 2004, she obtained her bachelor degree in Arts (B.A.) as a distinguished graduate. She was also selected as a member of "China Synergy Program for Outstanding Youth". Subsequently, she worked as a lecturer and student counselor in Sun Yat-sen University. In 2006, she received a fellowship for Excellent Young Lecturer and was elected as one of "the Best Student Counselor". From 2006-2007, she won a highly prestigious national scholarship to study linguistics (M.A.) in Leiden University. From 2008-2013, she worked on her PhD project in Cognitive Psychology Department, Leiden University. In 2010, she obtained a grant for academic expenditure from Leiden University Funds (LUF). From March, 2013 on, she will work as a post-doc researcher in Tilburg University.

个人简介

赖军，1982 年 8 月 1 日出生于中国赣州。2000 年，她毕业于赣州三中；同年以优异成绩考入中山大学，并获得为期四年的“凯思”优秀新生高额奖学金。2000-2004 年本科期间，她连续每年荣获校级奖学金。2004 年获得文学学士学位(英语语言与文学)，并荣获“中山大学优秀毕业生”殊荣；同年，参加“海外杰青汇中华”代表团，受到中央领导人和香港特别行政区长官接见。2004-2006 年，留校于中山大学，任外国语学院教育管理教师。2006 年，荣获“中山大学优秀辅导员”、“中山大学优秀团干”、“中山大学优秀外国语学院周宗哲奖教金”；同年，国家公派至荷兰莱顿大学攻读硕士学位。2007 年，获得文学硕士学位(语言学)。2008-2013 年 2 月，在莱顿大学攻读博士学位（心理学）。2010 年获得莱顿大学学术津贴。2013 年 3 月起，她将于蒂尔堡大学开始博士后工作，将任职于蒂尔堡逻辑与哲学中心、认知与交流中心。

List of Publications and Presentations

Publications

Lai, J., & Poletiek, F. (in press). How “small” is “starting small” for learning hierarchical structures? *Journal of Cognitive Psychology*.

Lai, J. & Poletiek, F.H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118(2)*, 265-273.

Lai, J., & Poletiek, F. (2010). The impact of starting small on the learnability of recursion. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 1387-1392)*. Austin, TX: Cognitive Science Society.

Poletiek, F. & Lai, J. (2012). How Semantic Biases in Simple Adjacencies affect Learning a Complex Structure with Non-Adjacencies in AGL. A Statistical Account. *Philosophical Transactions of the Royal Society B, 367,* 2046-2054.

Lai, J., & Poletiek, F. (submitted). A semantic-syntactic account of recursive sentence processing in Dutch.

Poletiek, F., Conway, C.M., Ellefson, M.R., Lai, J., & Christiansen, M.H. (submitted). Effects of starting small in artificial grammar learning of recursive structure.

Schiller, N. O., Lai, J., & Verdonschot, R.G. (submitted). Onset effects in a script without onsets: reading aloud and picture naming in Chinese.

Verdonschot, R.G., You, W.P., Lai, J., & Zhang, Q.F. (submitted). Priming in Chinese: syllabic or a matter of overlap?

Conference Presentations

6/2012	“Flexible” or “Strict” starting small effect? (Poster). EPOS Annual meeting. (Amsterdam)
12/2011	The facilitation of probabilistic input in processing recursion (Poster). NVP Winter Conference on Cognition, Brain, and Behavior of the Dutch Psychonomic Society. (Egmond aan Zee, the Netherlands)
11/2011	A probabilistic perspective on incremental learning of center-embedded (CE) structures (Poster). The 52nd Annual Meeting of the Psychonomic Society (Seattle, USA)
09/2011	Frequency distribution cues in processing recursive structures. (Poster). The 17th Meeting of the European Society for Cognitive Psychology. (ESCoP, Donostia-Sebastián, Spain).
10/2010	The influence of adjacent-dependencies and starting small on learning recursion (Oral presentation). The 32nd Annual Meeting of Cognitive Science Society (Portland, USA)
09/2009	The impact of starting small: the learnability of hierarchical structures in AGL (Poster). The 16th Meeting of the European Society for Cognitive Psychology (ESCoP, Krakow, Poland).