The learnability of center-embedded recursion : experimental studies with artificial and natural language

Lai, J.

Author: Lai, Jun

Title: The learnability of center-embedded recursion : experimental studies with artificial and natural language

Issue Date: 2013-02-26

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.