# Learning Recursion: Multiple Nested and Crossed Dependencies

Meinou H. de Vries, Morten H. Christiansen
& Karl Magnus Petersson

Language acquisition in both natural and artificial language learning settings crucially depends on extracting information from ordered sequences. A shared sequence learning mechanism is thus assumed to underlie both natural and artificial language learning. A growing body of empirical evidence is consistent with this hypothesis. By means of artificial language learning experiments, we may therefore gain more insight in this shared mechanism. In this paper, we review empirical evidence from artificial language learning and computational modeling studies, as well as natural language data, and suggest that there are two key factors that help determine processing complexity in sequence learning, and thus in natural language processing. We propose that the specific ordering of non-adjacent dependencies (i.e. nested or crossed), as well as the number of non-adjacent dependencies to be resolved simultaneously (i.e. two or three) are important factors in gaining more insight into the boundaries of human sequence learning; and thus, also in natural language processing. The implications for theories of linguistic competence are discussed.

*Keywords:*      artificial language learning; non-adjacent dependencies; processing complexity; recursion

## 1. Introduction

### 1.1. Competence versus Empirical Observations

One must not make too much of the exact form of the competence theory in the related task of building a broader psychological theory.

(Pylyshyn 1973: 45)

A theory of psychological processing typically focuses on actual and measurable performance. This is the perspective taken in the current paper with respect to structured-sequence processing in general as well as human language processing. From this point of view, it is natural to view the language faculty as a neurobiological system. The task, then, is to characterize the representational, processing and acquisition properties of this system at the neurobiological and psychological levels. In contrast, considerable work in theoretical linguistics, such as formal language theory, has focused on describing an idealized competence, comprising the knowledge of language that a speaker/hearer supposedly has. Instead of being grounded in experimental evidence as support, competence theories are mostly supported by linguistic intuitions (Pylyshyn 1973) and abstract computational considerations. Formal language theory might therefore not be the best source of information about the boundaries of human language processing.

One well-known intuition about syntactic structure is the property of recursion, an operation that permits a finite set of rules to generate an infinite number of expressions. Empirical evidence, however, has demonstrated that people are only able to generate and process recursive constructions to a very limited extent. Yet, linguists have concluded that recursion is a fundamental, possibly innate and unique part of the human language faculty.

One may ask whether we really need a competence theory that incorporates unbounded recursion (see e.g. Levelt 1974, Christiansen 1992, Petersson 2005). We stress that empirical observations about language processing mechanisms are more useful in the enterprise of understanding human language processing than linguistic intuitions. Thus, in this paper, the focus is on empirical observations from a diversity of experimental techniques (e.g., behavioral experiments, functional neuroimaging and computational modeling). More specifically, we concentrate on recursive structures involving multiple overlapping non-adjacent dependencies, the existence of which has been suggested by generative linguistics to be one of the major challenges for empirically-based approaches to language (Tallerman *et al.* 2009).

## 1.2.   *Non-Adjacency in Language*

Non-linear relationships between words are very characteristic of natural languages. For instance, in the sentence *The dog that scared the cat ran away*, we need to link *the dog* to the verb phrase further down the sentence, *ran away*, in order to understand that it was the dog that ran away. We refer to these non-linear relationships as 'non-adjacent dependencies' (as opposed to 'adjacent dependencies'), and they are inherent to the hierarchical nature of human language representations. It may be obvious that non-adjacency adds structural complexity to human language, and thereby processing complexity, but exactly *how* is still topic of discussion. In this review, we investigate two factors that help determine the processing consequences of such structural complexity in language: (i) the way in which non-adjacent dependencies are ordered, and (ii) the number of non-adjacent dependencies that need to be resolved simultaneously (i.e. keeping multiple elements active until they are linked to their co-dependent).

*1.2.1. The Ordering of Non-Adjacent Dependencies*

Across languages, non-adjacent dependencies may be instantiated in several different ways. One instantiation of non-adjacent dependencies involve *nested center-embedded* dependencies. Here, the dependencies are embedded within one another, exemplified in the structure $A_1A_2A_3B_3B_2B_1$, where $A_i$ is the element that needs to be linked to element $B_i$. In this paper, we will refer to this type of non-adjacent dependencies as *nested dependencies.* Another instantiation of non-adjacent dependencies involves *crossed-serial* dependencies, where the dependencies between elements cross each other, exemplified in the structure $A_1A_2A_3B_1B_2B_3$ which we will refer to as *crossed dependencies.* In Figure 1, we depict both types of non-adjacency. It also demonstrates that non-adjacent dependencies are indeed exhibited differentially across languages, in this case German and Dutch, which are otherwise closely related. Note that both crossed and nested orderings can only exist if the number of dependencies is more than one; in other words: The existence of *multiple* dependencies is a *sine qua non* condition for the occurrence of crossed and nested dependencies.
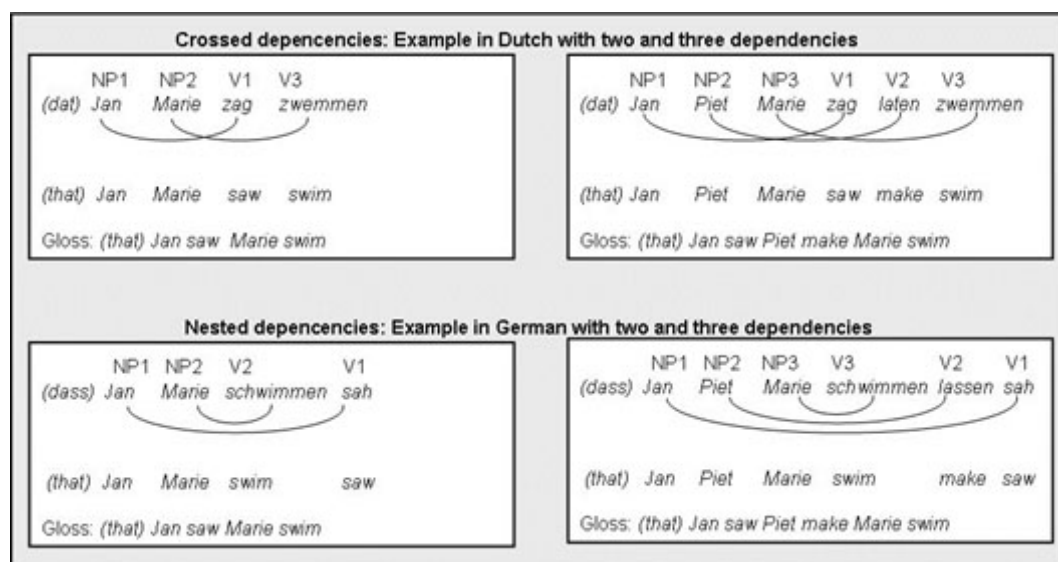


*Figure 1: Different ways of expressing non-adjacent dependencies in German and Dutch*

*1.2.2. Multiple Non-Adjacent Dependencies and the Intuition of Infinitude*

Figure 1 shows sentences with two and three dependencies (note that we refer to *dependencies*, not *embeddings* — a sentence with three dependencies contains two embeddings), respectively. In principle, one could keep on producing nested and crossed dependencies, and thus generating sentences of unbounded length. However, since humans possess finite brains that are constrained by (among other things) memory limitations, we have problems comprehending and producing sentences with three or more nested or crossed dependencies (e.g. Wang 1970, Hamilton & Deese 1971, Blaubergs & Braine 1974, Hakes *et al*. 1976, Bach *et al*. 1986). That is, people have difficulties keeping three or more elements active that are not yet linked to their co-dependents. Yet, the concept of infinite lingu-

istic competence has attracted much attention in theoretical linguistics since the 1950s. The mere existence of multiple crossed and nested dependencies may have led to an intuition of infinitude. Pullum & Scholz (2010) suggested that the notion of infinitude is due to researchers sticking to the mathematical notion that languages are sets: Since one can always think of a sentence that is longer than its precedent, and therefore the set of all sentences has to be infinite.

Infinitude then may be operationalized by the mathematical procedure of recursive definition, i.e. *recursion*. For example, an operation in which the same function is iteratively applied to its output (e.g., x → AxB and x → ∅ recursively defines ∅, AB, AABB, AAABBB, AAAABBBB, and so on, by rewriting or substitution). Both crossed and nested dependencies can be produced by unbounded, but also bounded, recursive operations. Indeed, allowing such an unbounded operation in theoretical models of natural language renders it infinitive, as it enables an infinitive number of possible sentences that can be created. However, the inference from actual syntactic phenomena observed in real sentences to the assumption of infinitude is not licensed (Pullum & Scholz 2010). Yet, the "standard argument" (terminology from Pullum & Scholz 2010) that grammars of natural language must contain recursive rule sets or recursive operators, is still prevalent among many linguists: The operation of recursion has often been portrayed as an essential and unique property of human language (Lasnik 2000, Hauser *et al*. 2002). For instance, Epstein & Hornstein (2004; cited in Pullum & Scholz 2010) stated the following:

> This property of discrete infinity characterizes *every* human language; none consists of a finite set of sentences. The unchanged central goal of linguistic theory over the last fifty years has been and remains to give a precise, formal characterization of this property and then to explain how humans develop (or grow) and use discretely infinite linguistic systems.
> (Epstein & Hornstein 2004; cited in Pullum & Scholz 2010: 113)

Why do so many linguists believe that grammars of natural language incorporate unbounded recursion, one way or another, in the absence of empirical evidence thereof? The inference that natural language grammars have unbounded recursive rules is based on a simplicity account (Lasnik 2000, Perfors *et al*. 2010). Indeed, a non-recursive grammar would be large if it were to generate a natural language. For example, it would require additional sets of rules for each additional depth of recursive expansion, and thus, any evaluation metric favouring shorter and simpler grammars should prefer a recursive grammar (Perfors *et al*. 2010). However, this is not true for neural networks (Siegelmann 1999), as was suggested in Elman (1991); see also Christiansen & Chater (1999). Here, each instantiation of a recursive construction is actually treated slightly different from each other, which is likely to be the case for sentence processing, as it unfolds in the human brain. Moreover, realistic neural networks have natural bounds on memory and processing precision (Petersson *et al*. 2010).

### 1.3.   *Infinitude and Empirical Data*

With the advent of generative grammar and recursion becoming key to achieving

discrete infinity (e.g. Chomsky 1956), early psycholinguistics devoted considerable effort to the study of nested dependencies (i.e. constructions observed in natural language, see above for our explanation of the terminology). After a brief hiatus, recursion is once again attracting attention as a hypothesized key feature of the language faculty, with the suggestion that recursion may be the only property of core language that is both species- and domain-specific (Hauser *et al*. 2002). Especially the case of nested dependencies, which will be the focus of our paper, has been thoroughly investigated, mainly through artificial language learning, and primarily with the presupposition that this paradigm taps into mechanisms hypothesized to be unique to humans (e.g. Fitch & Hauser 2004, Friederici 2004, Friederici *et al*. 2006), as will be discussed in further detail below. However, the empirical data do not match well with a grammar that contains unbounded recursion, in an important sense, as it would lead to serious overgeneralizations, stipulating very long sentences that are never used, and in fact, has never been observed (e.g. Christiansen 1992, Perfors *et al*. 2010). However, this is not a problem for bounded recursive procedures, or equivalent analogues (Petersson 2005, 2008). Indeed, soon after the advent of generative grammar, it was discovered that actual human performance on such constructions was at odds with the notion of infinite recursion. Recently, cross-linguistic studies have shown that unbounded recursion is not present in at least one natural language (see Everett 2005 for his work on the Pirahã language). Crucially, Pirahã is a fully fletched human communication system, with equal expressive power as in any other human language. As Everett (2005: 631) puts it, "Pirahã most certainly has the communicative resources to express clauses that in other languages are embedded." Thus, unbounded recursion is not a *necessary* component of any given language, and probably not of any processing account for human languages in general.

Furthermore, it was found that English sentences with more than two nested dependencies (see Figure 1 for an example of a sentence with three nested dependencies) are read with the same intonation as a list of random words (Miller 1962), cannot easily be memorized (Miller & Isard 1964, Foss & Cairns 1970), are difficult to paraphrase (Hakes & Foss 1970, Larkin & Burns 1977) and very difficult to comprehend (Wang 1970, Hamilton & Deese 1971, Blaubergs & Braine 1974, Hakes *et al*. 1976), and are judged to be ungrammatical (Marks 1968). Moreover, these limitations were soon discovered not to be unique to English but are also found in other European languages, such as German (Bach *et al*. 1986), French (Peterfalvi & Locatelli 1971), and Spanish (Hoover 1992) as well as in Hebrew (Schlesinger 1975), Japanese, and Korean (Uehara & Bradley 1996, Hagstrom & Rhee 1997). Only recently, Karlsson (2007) wrote an extensive review that illustrates how important "performance" is in the debate about unbounded recursion. From five major data sources from different languages, he extracted 119 sentences that contained multiple nested dependencies. From these, he concluded that the maximum number of nested dependencies was three (though this was very rare), and that in spoken language, multiple nested dependencies are practically absent. This suggests that "[f]ull-blown recursion creating multiple clausal center-embeddings is not a central design feature of language in use" (Karlsson 2007: 365).

We contend that it may be of greater importance to investigate our ability to process *certain types* of non-adjacent dependencies, such as nested and crossed dependencies. More specifically, we propose that the number (e.g., two or three) and ordering (e.g., embedded or nested) of these dependencies, as outlined above, might indicate where the empirical boundaries of human language processing lie. This is in line with Newport & Aslin (2004), who emphasized that the forms that non-adjacent dependencies take in natural language should be the focus of research:

> A learning mechanism additionally capable of computing and acquiring non-adjacent dependencies, while necessary for language learning, opens a computational Pandora's box: In order to find consistent non-adjacent regularities, such a device might have to keep track of the probabilities relating all the syllables one away, two away, three away, etc. If such a device were to keep track of regularities among many types of elements — syllables, features, phonemic segments, and the like — this problem grows exponentially. But, as noted, non-adjacent regularities in natural languages take only certain forms. The problem is finding just these forms and not becoming overwhelmed by the other possibilities.
>
> (Newport & Aslin 2004: 129)

In the next section, we review experimental work that has tested the learnability of non-adjacent dependencies in a laboratory-based *artificial language learning* setting, both in humans and non-human species.

## 2.      How Can We Test the Ability to Process Non-Adjacent Dependencies?

### 2.1.   *Mimicking Language Learning in the Lab*

One well-established way to test natural language phenomena in a laboratory-based setting, is using an artificial language learning (henceforth ALL) paradigm. Arthur Reber introduced this paradigm, and his early work was the first to focus on artificial grammar learning (AGL) tasks (Reber 1967, 1989). In the original task, subjects are asked to memorize a set of letter sequences generated by a finite state grammar, schematically displayed in Figure 2. Examples of valid letter sequences are MTTV, VXVRXRM, and MVRXRRM. After this memorization, the participants are told that the sequences that they just saw followed the rules of a grammar. They are then asked to classify a set of novel sequences as grammatical or ungrammatical, where half of these sequences obey the rules of the grammar whereas the other half does not. Typically, participants can perform this classification task with accuracy reliably above chance level, despite remaining largely unable to verbalize the exact rules of the grammar. Because of this dissociation between classification performance and the ability to explicitly describe the rules of the grammar, Reber classified this type of learning as implicit (Cleeremans *et al*. 1998).
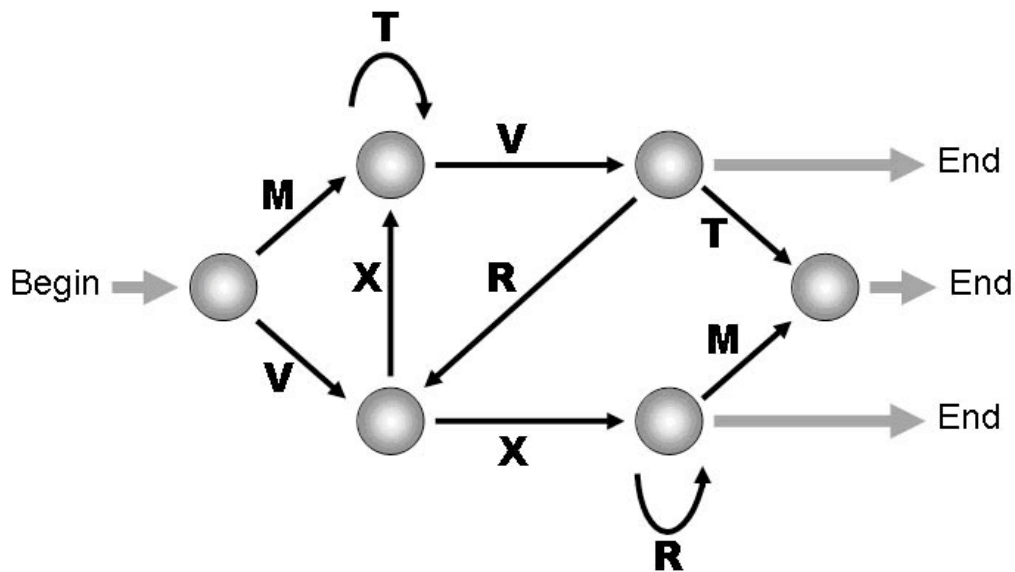
*Figure 2.  A finite-state grammar used to generate stimuli for artificial grammar learning tests (cf. Reber & Allen 1978).*

## 2.2.  Artificial and Natural Language Learning

The ALL paradigm has been employed widely to study different aspects of natural language learning though originally it was implemented to investigate the underlying *implicit sequence learning mechanism*, which is presumably shared with natural language learning (Reber 1967), as well as with other situations in which new skills have to be acquired. Indeed, skill learning crucially requires encoding, representing and production of structured sequences, and language is one excellent example of a domain where humans have to extract patterns from structured sequences in order to learn the underlying grammar (Conway & Pisoni 2008). The relation between language units, such as words, syllables and morphemes, adhere to certain sequence structures typical of language, of which crossed and nested dependencies are two examples. Determining how humans extract and use structural information from the environment is a great challenge for the cognitive neurosciences (Conway & Pisoni 2008), as is establishing the underlying neurobiological mechanisms of implicit sequence learning that mediates the acquisition of novel skills.

The neural correlates of implicit sequence learning as assessed by the AGL paradigm have been investigated by means of functional neuroimaging (e.g. Lieberman *et al*. 2004, Petersson *et al*. 2004, Forkstam *et al*. 2006; for an overview, see Petersson *et al*. 2004), brain stimulation (Uddén *et al*. 2008, de Vries *et al*. 2010), and special populations, such as Parkinson's Disease patients (e.g. Knowlton & Squire 1996, Reber & Squire 1999), participants diagnosed with Autism Spectrum Disorders (Brown *et al*. 2010), agrammatic aphasics (Christiansen *et al*. 2010), and dyslexics (e.g. Rüsseler *et al*. 2006, Pavlidou *et al*. 2009; for a review, see Folia *et al*. 2008), and generally involve frontal-striatal-cerebellar regions (Packard & Knowl-

ton 2002, Ullman 2004, but note the different terminology in the studies where implicit learning is sometimes referred to as procedural learning, and vice versa), which are also involved in the acquisition of grammatical regularities (Ullman 2004). More specifically, recent functional neuroimaging (e.g., Lieberman *et al.* 2004, Petersson *et al.* 2004, Forkstam *et al.* 2006) and brain stimulation research (Uddén *et al.* 2008, de Vries *et al.* 2010), implementing experiments based on the Reber paradigm, have identified which brain regions are involved in such a task. They have repeatedly shown that Broca's region, an area in the brain involved in syntactic processing of natural language, is also involved in artificial grammar processing. Indeed, breakdown of syntactic processing in agrammatic aphasia is associated with impairments in AGL (Christiansen *et al.* 2010). This supports the hypothesis that AGL taps into implicit sequence learning, and thus provides a useful way to investigate natural language processing (cf. Petersson *et al.* 2004).

The underlying implicit sequence learning mechanism appears to be rather domain-general, with evidence of learning in several domains (e.g., speech-like stimuli, tone sequences, visual scenes, geometric shapes, visuomotor sequences — see Conway & Pisoni 2008 for an overview). Conway & Pisoni investigated how implicit sequence learning in different domains contributes to language processing by directly linking performance of individual participants on a non-linguistic implicit sequence learning task to performance on a spoken sentence perception task, in which participants had to predict the final word of each sentence in low and high predictability conditions. They found that, indeed, individual variability in implicit sequence learning correlated with language processing. Supportive evidence also comes from a recent study by Misyak *et al.* (2010a, 2010b), who found that individual differences in learning non-adjacent dependencies, assessed by a non-linguistic implicit sequence learning task, strongly correlate with the processing of natural language sentences containing complex non-adjacent dependencies.

In sum, there is substantial evidence that language acquisition and language processing in both natural and artificial settings is mediated by a more general implicit sequence learning mechanism. By implementing ALL experiments, we thus tap into the underlying sequence learning mechanism that also mediates natural language acquisition, and the resulting processing system. Investigating the boundaries of this mechanism will therefore add to our understanding of human language acquisition and processing. For example, the empirical finding that we cannot understand sentences with more than three dependencies, for instance, is in accord with the fact that, as yet, no study of ALL has convincingly demonstrated that humans are able to do so in a well-controlled ALL setting (we will discuss this in more detail below).

In a review, Gomez & Gerken (2000) emphasize that one of the beneficial aspects of employing ALL paradigms, is that researchers obtain control over the input to which learners are exposed, and that it also controls for prior learning. Knowing what participants *can* learn, may then lead to more specific hypotheses about the actual mechanisms involved. Gomez & Gerken identified four aspects of language that successfully have been investigated through ALL tasks, in studies involving both infants and adults: Word segmentation, encoding and re-membering the order in which words occur in sentences, generalization of gram-

matical relations, and learning syntactic categories. Learning non-adjacent dependencies is yet another aspect that can, and is, being tested through ALL tasks. The usefulness of ALL paradigms is that one can design experiments capturing such key features of language. According to Gomez & Gerken, if we can isolate a specific linguistic phenomenon experimentally, we can go on to test it using a set of various manipulations. These manipulations are driven by our knowledge of natural language acquisition. Ultimately, the proof of the ALL approach will depend on the extent to which it generates new ways of understanding the mechanisms of natural language acquisition (Gomez & Gerken 2000).

### 2.3.    *Potential Pitfalls in Artificial Language Learning Settings*

Before discussing how non-adjacent dependencies may be tested in a laboratory setting, we would like to stress a few potential pitfalls in ALL settings. As mentioned in the previous paragraph, a common assumption in ALL research is that when participants are exposed to sequences generated by a specific grammar and subsequently are able to distinguish new grammatical items from ungrammatical ones, then participants have in some sense "learned" the structure of the underlying grammar. This notion goes back to Reber's original work, in which he suggested that his participants were "learning to respond to the general grammatical nature of the stimuli" (Reber 1967: 855).

More generally, the current tendency is to (implicitly) assume that if the sequences were generated by a particular type of grammar, for instance a phrase-structure grammar, and if participants show evidence of learning, then they have learned this particular phrase structure grammar and process the sequences according to the phrase-structure rules (e.g., Saffran 2001, Fitch & Hauser 2004, Thompson & Newport 2007, Makuuchi *et al*. 2009). Thus, strong claims are made about the formal properties of the regularities being learned even though performance in many of these experiments is only about 70% correct in terms of classifying novel items as grammatical or ungrammatical. However, none of these studies actually seek to determine whether the minimal computational machinery needed to account for the observed level of performance necessarily requires such a formalization of the knowledge in order to account for the results. In the absence of such explicit computational accounts of the experimental results, it is unclear whether such strong claims about the formal properties of the acquired knowledge are warranted. In other words, just because an experimenter uses a particular grammar formalism to generate the sequences to be learned, it does not necessarily mean that participants may not utilize a different, and perhaps much simpler, way of representing the knowledge acquired. As will be clarified in the next few paragraphs, this potential pitfall is a common mistake among experimenters using ALL paradigms, leading to over-interpretation of their results. A similar argument has very recently been put forward by Lobina (2011), specifically with respect to the recent ALL studies investigating recursion as a property of natural language. Lobina emphasizes that it is a common error in the ALL field to extrapolate to recursive parsing operation from the correct processing of structures that contain nested dependencies.

## 2.4. Learning Non-Adjacent Dependencies in the Lab

Whereas the learning of adjacent dependencies has been shown in laboratory-based settings repeatedly, both with visual, auditory and tactile stimuli, linguistic and non-linguistic material, and in infants, adults and non-human species (e.g. Saffran *et al.* 1996, 1999, Aslin *et al.* 1998, Hauser *et al.* 2001, Conway & Christiansen 2005, Perruchet & Pacton 2006, Forkstam *et al.* 2008, Gebhart *et al.* 2009), the learning of *non-adjacent* dependencies seems to be harder (Gebhart *et al.* 2009), though certainly possible to a certain extent. Gomez (2002) for instance, showed that the degree to which non-adjacent dependencies are learned depends on the relative variability of the intervening material (i.e. X in the pattern AXB, where A and B belong together), in both adults and 18-month-old infants. When X is highly invariant, that is, when X is drawn from a pool of only two alternatives, it is harder to learn the dependency between A and B than when X is drawn from a pool of 24 alternatives. In other words, the relationship between A and B stands out most when X is varied to a great degree, while keeping A and B relatively invariant. In contrast, Newport & Aslin (2004) and Onnis *et al.* (2005) found that the crucial factor for learning non-adjacent dependencies is rather the similarity between A and B (Perruchet & Pacton 2006).

However, different from the above mentioned studies, where participants learned to solve only one non-adjacent dependency at a time, the focus here is specifically on multiple overlapping non-adjacent dependencies, including nested dependencies, as depicted in Figure 1 for natural language, which requires *more than one* non-adjacent dependency being managed simultaneously. We will also discuss current findings on crossed dependencies (Figure 1), though to a limited extend, as findings on this type of structure are yet scarce. But we start by discussing recent experimental findings in both humans and non-human species that gave rise to such lively debate in the field.

## 2.5. Can Animals Handle Non-Adjacent Dependencies?

The finding that learning non-adjacent dependencies is considerably harder than learning adjacent dependencies has raised questions regarding the uniqueness of non-adjacent dependencies to human language processing. Hauser *et al.* (2001) had shown that *adjacent* dependencies are learnable by non-human primates (see also Heimbauer *et al.* 2010). Would this also be the case for non-adjacent dependencies?

### 2.5.1. Animals in Artificial Language Learning Situations

Newport *et al.* (2004) indeed showed that, in an ALL setting, non-human primates (New World monkeys) are capable of tracking *simple* non-adjacent dependencies, in situations where only one non-adjacent dependency needs to be resolved at a time. Given our assumption that there are two factors that determine processing complexity, namely (1) the number of non-adjacent dependencies that need to be resolved simultaneously, and (2) the ordering of these dependencies, a more relevant question is whether non-human species can resolve

nested and crossed dependencies (implying processing multiple dependencies simultaneously). Indeed, Fitch & Hauser (2004) showed that cotton-top tamarin monkeys, after a short training period, fail to learn structures that exhibit nested dependencies, though in their paper, dependencies were not indexed, such that simpler strategies could have been used to solve the test for nested dependencies (see Perruchet & Rey 2005 and de Vries *et al*. 2008 for criticism). A recent study by Gentner *et al*. (2006) claimed that song birds could learn such nested dependencies, after extensive training. However, also here, the dependencies between the elements were not indexed, such that simpler strategies could have been used by the birds to solve the task (see Corballis 2007a and de Vries *et al*. 2008 for criticism). In a recent experiment (van Heijningen *et al*. 2009), zebra finches were tested for their ability to classify nested dependencies. Interestingly, one zebra finch (out of eight) was able to generalize the acquired syntactic structure to another stimulus set. However, additional testing showed that no strategy was involved that required processing nested dependencies. Also here, a simpler strategy was used to solve the task (van Heijningen *et al*. 2009). Thus, as yet, no non-human species has been shown to be able to learn nested dependencies, due to methodological flaws (as argued by Perruchet & Rey 2005, de Vries *et al*. 2008, Corballis 2007a; see also Liberman 2004a, 2004b, Hochmann *et al*. 2008).

## 2.5.2. Non-Adjacent Nested Dependencies in a Natural Setting

The results of these ALL experiments in non-human animals parallel those found in studies looking at nested organization in the *natural behavior* produced by non-human primates, which possibly indicates to what extent nested dependencies are exhibited in non-human primates (Conway & Christiansen 2001). Two interesting studies describe the way in which capuchin monkeys, chimpanzees, bonobos (Johnson-Pynn *et al*. 1999), and human children (Greenfield *et al*. 1972) use strategies to combine cups, each varying in size such that the smallest cup could fit into the one that was larger, which in turn could fit into the next largest, and so on. When instructed or encouraged to nest the cups, only human children older than 20 months were able to use a nesting strategy, in which two or more cups are combined to form a single unit, which is then placed into another cup. Interestingly, the development of the cup-nesting strategy in children has parallels to the structural development of grammar and phonology in language (Greenfield 1991). The primates, however, were limited in their ability to perform the nesting cup task and did not utilize the complex nesting strategy, in which units are embedded within other units, but only adopted simpler strategies (Johnson-Pynn *et al*. 1999, Conway & Christiansen 2001).

Summarizing the above findings, it seems that non-human animals are able to track simple non-adjacent dependencies (as was shown in New World monkeys in Newport *et al*. 2004), but processing multiple non-adjacent dependencies that are embedded within one another may be beyond even our nearest primate cousins. This suggests, perhaps, that processing *multiple* non-adjacent dependencies simultaneously may be a specific human ability. The question whether this restriction holds for crossed and/or nested dependencies cannot be answered, as no study in the literature so far has looked at non-human ability to

process *crossed* dependencies. In conclusion, the number of non-adjacent dependencies that need to be resolved simultaneously may be a decisive factor in determining what is learnable to non-human species and what is not.

### 2.6. *Processing Nested Dependencies in Humans*

As in animal studies, the processing of nested dependencies in humans has been studied extensively, whereas only a few studies have focused on the processing of crossed dependencies. In the current section, we will discuss experimental findings regarding the processing of nested dependencies in humans.

Following up on the study of Fitch & Hauser (2004), Friederici *et al*. (2006) implemented a similar paradigm in an FMRI study. They set out to test the neural correlates of processing nested dependencies in humans. Also here, an ALL task was used. They found that the processing of sequences containing nested dependencies activated Broca's area (BA44/45). However, participants may have distinguished grammatical from ungrammatical sequences by merely counting the number of A and B elements and checking that they matched (or not), which was referred to as a counting strategy (de Vries *et al*. 2008; see also Corballis 2007b). In other words, participants in Friederici *et al*.'s (2006) study were not required to resolve the nested dependencies, as was experimentally shown by de Vries *et al*. (2008). Comparing performance on testing situations without ruling out whether other strategies could have been applied is a common issue. ALL researchers should thus design their tasks carefully to be sure that participants cannot solve the task through strategies such as counting, repetition monitoring, or simply detecting an additional element that lacks a co-dependent in the sequence (as was the case in one of the violation types of Bahlmann *et al*. 2008). Thus, in order to be able to examine the basis for the classification performance, a careful design is required.

Although we do not doubt that humans *can* process nested dependencies in natural language (although to limited extent), it is difficult to mimic this in a laboratory-based setting. One potential way to ensure that participants learn nested dependencies, is to add perceptual cues to the elements that belong together, in order to promote learning the dependencies of interest (e.g., Müller *et al*. 2010); however, explicit problem solving may become involved as soon as the dependencies stand out too much. Nonetheless, Uddén *et al*. (2009) showed that implicitly learning nested dependencies might well be a matter of time: Uddén *et al*. successively trained their participants for nine days in a row with no evidence for explicit awareness of the relevant dependencies or the use of explicit strategies. Another way to improve learnability of nested dependencies has been demonstrated by Conway *et al*. (2003), who used a training paradigm that started with simpler constructions, followed by gradual increases in depth of recursive structure.

One possibility to avoid explicit problem solving when learning nested dependencies, may be implementing a serial reaction time (SRT) task involving nested dependencies, as has been done for simple non-adjacent dependencies (Misyak *et al*. 2010a, 2010b). Forthcoming results from our group show learning of both nested and crossed sequences using this paradigm. Remarkably, crossed

dependencies are learned better and faster than nested dependencies, both by German and Dutch participants, despite the fact that from the point of unbounded recursion, the former requires context sensitive grammars and the latter only context-free grammars. Just the opposite of what might have been predicted based on the Chomsky hierarchy. This underlines the conclusion of Bach *et al*. (1986), who, in a psycholinguistic study, found that processing three crossed dependencies in Dutch was relatively easier than processing three center-embedded dependencies in German (see also Figure 1). The advantage of crossed over nested dependencies disappears when the number of dependencies that need to be resolved is reduced to two.

### 2.7.  *The Difference between Two and Three*

Based on the findings discussed above, our hypothesis is that, in sequence learning, and potentially also for natural language, the ordering of non-adjacent dependencies (crossed or nested) is an important factor only when there are three (or more) dependencies. In the case of two non-adjacent dependencies that need to be resolved simultaneously, there is no apparent difference in processing complexity between the nested or crossed ordering. In other words, the first factor that determines the demands on memory, and hence processing complexity, is the number of dependencies that need to be resolved simultaneously. If that number exceeds two, then the factor "ordering" becomes important. Indeed, this is supported by the natural language findings of Bach *et al*. (1986), demonstrating that differences in processing difficulty between Dutch and German is only present when the number of dependencies exceeds two. Preliminary data from our group further support this prediction from a more domain-general sequence learning perspective, using the combined AGL–SRT paradigm mentioned above (Misyak *et al*. 2010a, 2010b). Figure 3 provides a schematic overview of the suggested complexity levels (from a processing perspective).
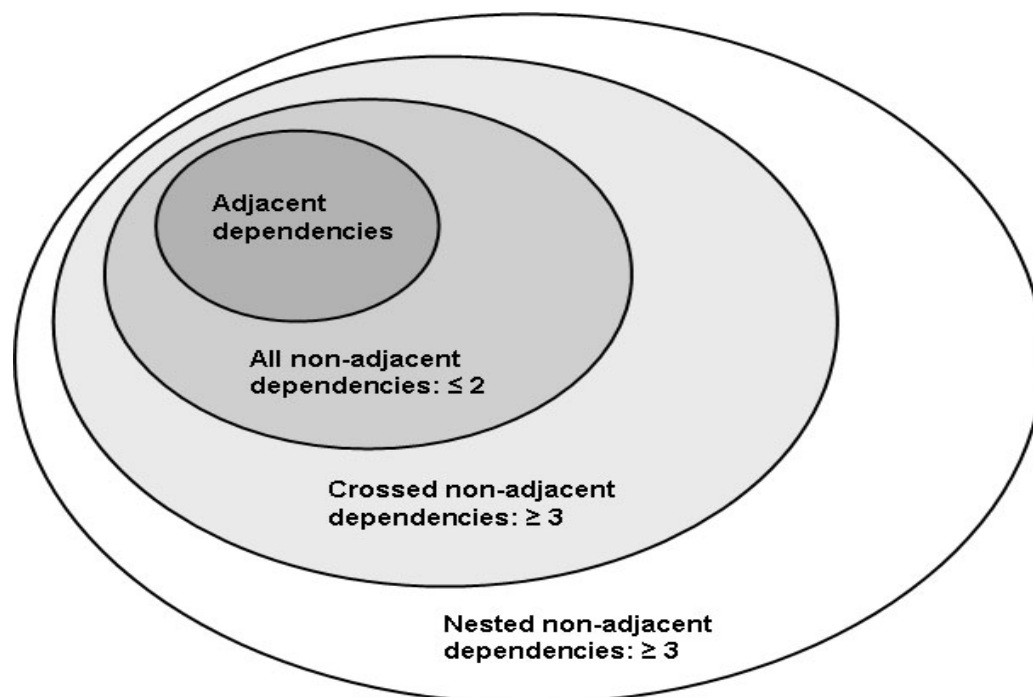
*Figure 3: A schematic overview of our suggested levels of processing complexity. Note that each level of this hierarchy denotes a decisive factor that adds to processing complexity. Thus, it is not the case that three nested dependencies are harder to process than, say, six crossed dependencies. Instead, this figure emphasizes that the presence of more than two dependencies is a* sine qua non *condition for measurable differences in processing complexity between nested and crossed dependencies.*

The consequence of this reasoning is that the traditional differences between context-free and context-sensitive grammars, as put forward by the Chomsky hierarchy, is less relevant for understanding the language system of the human brain (see also Uddén *et al.* 2009, Petersson *et al.* 2010). This insight is of potential importance to the many ALL researchers who view the Chomsky hierarchy as uniquely informative about the human language faculty, and subsequently base their experiments on this assumption, for instance by directly comparing acquisition performance on certain sequences with grammars from different levels of the Chomsky hierarchy (e.g. Fitch & Hauser 2004, Friederici *et al.* 2006; for similar criticism, see Lobina 2011). Subsequently, many of these researchers draw conclusions about the underlying knowledge structures (i.e. 'competence', e.g. Fitch & Hauser 2004) or operational processes ('hierarchical processing', e.g. Friederici *et al.* 2006). Instead, we suggest that the way forward is to focus on processing complexity and the different levels of complexity that *sequences* may take.

Very little work has been done on the processing of crossed dependencies, specifically in the field of ALL (see Uddén *et al.* 2009 for an exception). Yet, there are several arguments that support our hypothesis that crossed dependencies are easier than nested dependencies, if the number of dependencies exceeds two. Below, we will briefly discuss evidence from cross-linguistic psycholinguistic

experiments, computational simulations, and ALL studies.

### 2.7.1. Cross-Linguistic and Psycholinguistic Support

The only study, that directly investigated complexity differences between crossed and nested dependencies in natural language processing, is that of Bach *et al.* (1986). They asked native German speakers to provide comprehensibility ratings of German sentences containing nested dependencies and native Dutch speakers to rate Dutch sentences containing crossed dependencies (examples are depicted in Figure 1). They found no difference in processing difficulty between crossed and nested structures when *two* dependencies had to be resolved. However, when sentences contained *three* dependencies, nested dependencies (in German) were harder to process than crossed (in Dutch).

### 2.7.2. Support from Computational Simulation

Christiansen & MacDonald (2009) modeled the comparative difficulty of nested versus crossed dependencies by training a Simple Recurrent Network (SRN; Elman 1990) on sentences containing such dependencies. Their simulation results demonstrated that the SRNs exhibited the same pattern of processing difficulties as humans: Crossed dependencies were found easier than nested, but only when there were three dependencies. When there were two dependencies, no qualitative differences were found (see also Christiansen & Chater 1999 for similar results with simpler languages more akin those used in ALL).

### 2.7.3. Support from Artificial Language Learning Settings

Uddén *et al.* (2009) showed in an ALL study that Dutch participants performed better on crossed than on nested dependencies. They implemented an implicit AGL paradigm, extending the acquisition phase to nine days in a row for each participant. Their results suggested that successful performance on the two grammar types differed most for the longer test sequences with three dependencies, although this difference did not reach significance. The question whether the better performance on crossed dependencies is due to the participants being Dutch, and hence, familiar with such structure in their native language, or if crossed dependencies are intrinsically easier to process is not answered in this study. However, forthcoming results from our group showed that, in a combined AGL–SRT study, learning crossed dependencies is easier than nested dependencies, both in German and Dutch participants. We suggest that future research should focus not only on the processing differences between crossed and nested dependencies, but specifically on the processing differences between sequences with two non-adjacent dependencies and three (or more) non-adjacent dependencies, both in crossed and nested order.

### 2.7.4. Support from the Starting Small Principle

More indirect support for our hypothesis that crossed dependencies are easier to

process than nested dependencies (if the number of dependencies exceeds two) comes from a study looking at the "starting small" principle (Elman 1993). Conway *et al*. (2003) showed that participants, who were being trained on nested dependencies, learn better if they are exposed gradually to an increased number of dependencies. Participants were first exposed to short sequences with only one dependency relation, then to sequences with two dependencies, followed by three dependencies. Although the benefit of starting small has not been shown for the acquisition of *crossed* dependencies (for which the effect may be smaller assuming that crossed dependencies are easier to learn than nested dependencies), this highlights toward the importance of the *number* of dependencies that need to be processed simultaneously. Again, this is not accounted for in terms of the Chomsky hierarchy.

### 2.7.5. Support from the Missing Verb Effect

The importance of distinguishing between two versus three (or more) dependencies is further underscored by studies on the so-called "the missing verb effect". Gibson & Thomas (1999) investigated the role of memory limitations in the processing of sentences that contained three nested dependencies. They found that when deleting the second VP in a sentence (*'was cleaning every week'* in (1a)), the resulting ungrammatical sentence (1b) was rated just as acceptable as the original grammatical version in an off-line rating task. This was argued to be caused by working memory saturation.

(1)   a.   The apartment that the maid who the service had sent over was cleaning every week was well decorated.
      b.   * The apartment that the maid who the service had sent over was well decorated.

Testing predictions from a neural network model, Christiansen & Mac-Donald (2009) conducted an on-line sentence processing study with the same materials and found that the ungrammatical (1b) was actually rated better than the grammatical (1a). They replicated these result using materials controlled for length and semantic plausibility. Interestingly, these were all native English participants, where nested dependencies are relatively infrequent. The missing verb effect has also been replicated in French (Gimenes *et al*. 2009). In this study, the effect was reduced when the third noun phrase was replaced by a pronoun, making the reader more sensitive to the missing second VP. Vasishth *et al*. (2010) conducted a similar study with German participants and found that they were not sensitive to the missing verb effect as illustrated in (1b). They suggested that this difference was caused by the participants' adaptation to the specific grammatical properties of German: In contrast to English, German subordinate clauses always have the verb in clause-final position. Hence, the German speakers may maintain predictions about upcoming sentence parts more robustly compared to English speakers. This again shows that there are critical processing differences between two or three non-adjacent dependencies, although the German case may be exceptional. An interesting question is whether crossed

dependencies also exhibit the missing verb effect. Given our assumption that crossed dependencies are easier to process than nested, these structures may be less prone to the missing verb effect. In line with this, preliminary AGL–SRT results from our group suggest that this is indeed the case: It appears that the missing verb effect can be replicated in nested, but not in crossed dependencies, for both German and Dutch participants.

In conclusion, the number of dependencies that need to be processed simultaneously is an important factor in determining processing complexity. Given existing results, this difference is seen already between two and three dependencies. Natural and artificial language results (Bach *et al*. 1986, Uddén *et al*. 2009) and computational modeling results (Christiansen & Chater 1999, Christiansen & MacDonald 2009) support the suggestion that, when the number of dependencies is two or less, there is no difference in processing cost between crossed and nested structures. When the number of dependencies exceeds two, crossed dependencies are found easier to process than nested. Further studies are needed to precisely establish the cross-linguistic support for this suggestion.

## 3.      The Neural Correlates of Processing Non-Adjacent Dependencies

Several functional neuroimaging studies have compared the processing of sequences containing non-adjacent dependencies with sequences containing adjacent dependencies (Friederici *et al*. 2006, Bahlmann *et al*. 2008) and the results show that Broca's region is relatively more engaged in processing sequences containing non-adjacent dependencies. However, Broca's region is also engaged in the processing sequences generated from a simple right-linear grammars (Petersson *et al*. 2004, Forkstam *et al*. 2006, Petersson *et al*. 2010).

### 3.1.   *Working Memory and Non-Adjacency*

These findings are not surprising, given that nested dependencies owe their complexity to the fact that they cannot be resolved immediately: The first element has to be kept activated until its referent is encountered; hence, short-term memory remains loaded. This is not the case in adjacent dependency resolution, where an element can be discharged right away without encountering intervening material. Interestingly, in a simple working memory task (0-back, 1-back, 2-back, 3-back), Braver *et al*. (1997) showed exactly this: The activation level in Broca's region increased as a linear function of the distance between the element and its co-dependent (in this case, detecting repetitions in the *n*-back task). Activation of Broca's region as a result of the comparison between non-adjacent and adjacent dependencies could therefore very plausibly have been caused by differences in memory load between the two tasks (see also de Vries *et al*. 2008 for a similar suggestion). After all, matching syllables (as was involved in the tasks by Friederici *et al*. 2006, Bahlmann *et al*. 2008) presumably is not so much different from matching letters (Braver *et al*. 1997), irrespective of whether test sequences are generated by an artificial language or by an n-back task. The relative processing complexity of sequences containing nested dependencies may

therefore be directly related to memory load. However, it is likely that to regard differences in memory load as the differentiating factor between processing non-adjacent and adjacent dependencies is too simplistic — and not only because the notion of working memory is still not settled upon. In contrast, it is hard to draw a sharp line, on theoretical grounds, between on-line processing memory and representational processing itself (Minsky 1967). Instead, we want to emphasize that there are presumably limitations on the on-line sequence memory available for structured sequence processing that are determined by neurobiological factors, and possibly also linguistic experience, such that experience with a specific language might affect the ease with which multiple non-adjacent dependencies are resolved (see also Christiansen & MacDonald 2009 for further discussion). Future research should elaborate on this possibility.

### 3.2. *Disentangling Memory Effects and Complexity*

In an attempt to segregate syntactic complexity and memory effects in *natural* language processing, Makuuchi *et al*. (2009) implemented an event-related fMRI study and found that distance between syntactic elements and whether or not a sentence contains nested dependencies are two separate factors. The former involved the left inferior frontal sulcus, and the latter the left pars opercularis — the posterior part of Broca's region. Petersson *et al*. (2010), however, have suggested that these sub-regions are too close in space to reliably resolve with standard fMRI.

Makuuchi *et al*. (2009) compared four experimental conditions that contained natural language sentences of different forms: (1) Hierarchy and Long Distance, (2) Hierarchy and Short Distance, (3) Linear and Long Distance, and (4) Linear and Short Distance. A potential weakness in this manipulation however, may be that the difference between 'Hierarchy' and 'Linear' was employed such that in the Hierarchy conditions, more than one non-adjacent dependency needed to be resolved, whereas in the Linear condition, there was only one (despite referring to this condition as 'linear', the crucial elements in those sentences still had to be linked to elements further away in the sentence). Thus, the conditions used in the study rather exemplified situations where one versus multiple non-adjacent dependencies has to be established. Although the authors claim to have segregated memory load from structural complexity, this logically cannot be the case: Establishing multiple non-adjacent dependencies simultaneously *must* entail more memory load than establishing only one. Furthermore, it also shows that disentangling structural complexity from memory is difficult, if at all possible. Moreover, Petersson *et al*. (2010) show that the sub-region identified by Makuuchi *et al*. (2009) as engaged in the processing of sequences with nested non-adjacent dependencies, is also engaged in the processing of simple right-linear structures, where there are no requirements to process hierarchically nested non-adjacent dependencies at all. Rather, we think that processing complexity is intrinsically tied to the memory resources required, and likely also relevant processing experience. Thus, our suggested complexity levels are determined in part by the intrinsic memory constraints of the underlying sequence learning mechanism. Finally, it is not clear from Makuuchi *et al*.'s study, which

employed natural language material, that the reported differences are related to sentence-level syntax and not for example sentence-level semantics. In normal language processing, semantics, phonology and syntax operate in close spatial and temporal contiguity in the human brain. Therefore, the AGL paradigm has been used to create a relatively uncontaminated window onto the neurobiology of syntax (Petersson *et al*. 2004, 2010).

### 3.3. Complexity and Broca's Region

The question remains as to why the ALL studies by Petersson *et al*. (2004) and Forkstam *et al*. (2006), also using an event-related fMRI design, revealed a firmly replicated (Uddén *et al*. 2008, Petersson *et al*. 2010) activation in Broca's region during the processing of sequences generated from simple right-linear grammars. In these studies, there was no comparison between conditions that contained adjacent versus non-adjacent dependencies, which was the case in the studies reported by Friederici *et al*. (2006) and Bahlmann *et al*. (2008). Instead, processing of adjacent dependencies was compared against a sensorimotor decision baseline. Lacking a condition that directly compared adjacent versus non-adjacent dependencies, only one conclusion can be drawn, namely, that the activation of Broca's region is *not* specific to structures that entail non-adjacent dependencies.

In sum, the mere presence of non-adjacent dependencies adds to processing complexity, as does the *number* of to-be-established non-adjacent dependencies. Activation in Broca's region however cannot be specific to those situations only, as is pointed out in Petersson *et al*. (2010). Rather these findings, in conjunction with functional neuroimaging data from other domains requiring sequence processing (for reviews see e.g., Petersson *et al*. 2004, 2010), suggest that Broca's region is a generic on-line structured sequence processor that is activated at different levels depending on processing complexity.

### 4.    Implications for Theories of Natural Language Processing

A significant and growing body of experimental evidence from a range of experimental approaches (behavioral experimentation, functional neuroimaging, brain stimulation, brain lesion studies, computational modeling, etc.) reviewed here converge on the suggestion that natural and artificial language processing share underlying sequence learning mechanism(s). By conducting ALL experiments, one of the aims is to tap into this mechanism, providing additional insights in the boundaries on structured sequence processing, in general, and natural language acquisition and processing, more specifically.

When we acquire language (or other skills dependent on structured sequence processing), we need to extract regularities from input that is sequential in nature. Regularities exist when elements are linked in specific situations. Thus, identifying dependencies between input elements is a way to systematize input, and in conjunction with prior domain-general and domain-specific constraints, induce models for generalization. It is natural to suppose that we are constrained by memory limitations; and thus we can only extract certain patterns from the in-

put implicitly. We think one important use of ALL paradigms is to explicitly characterize these boundaries on cognition in order to provide a better understanding of the cognitive mechanisms that enable us to extract regularities from the input, including natural language acquisition.

We agree with Lobina's (2011) argument that a common error of ALL experiments is, as we ourselves have pointed out above, that sequence processing is often mistaken to be uniquely informative of purported underlying formal parsing operations. A crucial question is whether the distinction between competence and performance is helpful at this stage of scientific inquiry. We suggest that our results speak to how processing and knowledge of language are fundamentally intertwined in a way not well-captured by traditional approaches in formal language theory. Crucially, though, our proposed levels of processing complexity in Figure 3 should not be interpreted as indicating that the human language processing system favors context-sensitive over context-free competence grammars. Indeed, these concepts (and the Chomsky hierarchy from which they are derived) are orthogonal to the points we make. Instead, our focus is on performance. We suggest that the sequence processing system — and thus the cognitive processes that depend on it, such as natural language — may be constrained by (i) the number of dependencies that need to be resolved, and (ii) the ordering of these dependencies. The difference between nested and crossed dependencies becomes relevant only when the number of dependencies exceeds two. This may not be specific to the language domain, but a domain-general constraint; given that many ALL studies have been shown to be replicable with stimuli in different domains and modalities (though some differences do exist; e.g. Conway & Christiansen 2005).

To conclude, we have argued that processing complexity relating to structured sequence processing may be determined by (i) the number of dependencies that need to be resolved, and (ii) the ordering of these dependencies. Considering this assumption as a point of departure, several new research questions can be explored. To do so, artificial language learning paradigms may be implemented to explore the boundaries of the sequence learning mechanism shared with natural language.

## References

Aslin, Richard N., Jenny R. Saffran & Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9, 321–324.

Bach, Emmon, Colin Brown & William Marslen-Wilson. 1986. Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1, 249–262.

Bahlmann, Jörg, Ricarda I. Schubotz & Angela D. Friederici. 2008. Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534.

Blaubergs, Maija S. & Martin D. Braine. 1974. Short-term memory limitations on

decoding self-embedded sentences. *Journal of Experimental Psychology* 102, 745–748.

Brown, Jamie, Aczel Balzs, Luis Jiménez, Scott B. Kaufman & Kate P. Grant. 2010. Intact implicit learning in autism spectrum conditions. *The Quarterly Journal of Experimental Psychology* 63, 1789–1812.

Braver, Todd S., Jonathan D. Cohen, Leigh E. Nystrom, John Jonides, Edward Smith & Douglas Noll. 1997. A parametric study of prefrontal cortex involvement of human working memory. *Neuroimage* 5, 49–62.

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

Christiansen, Morten H. 1992. The (non)necessity of recursion in natural language processing. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, 665–670.

Christiansen, Morten H. & Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23, 157–205.

Christiansen, Morten H., M. Louise Kelly, Richard Shillcock & Katie Greenfield. 2010. Impaired artificial grammar learning in agrammatism. *Cognition* 116, 382–393.

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

Cleeremans, Axel, Arnaud Destrebecqz & Maud Boyer 1998. Implicit learning: News from the front. *Trends in Cognitive Sciences* 2, 406–416.

Conway, Christopher M. & Morten H. Christiansen. 2001. Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5, 539–546.

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

Conway, Christopher M., Michelle R. Ellefson & Morten H. Christiansen. 2003. When less is less and when less is more: Starting small with staged input. *Proceedings of the 25th Annual Conference of the Cognitive Science Society* 270–275.

Conway, Christopher N. & David B. Pisoni. 2008. Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences* 1145, 113–131.

Corballis, Michael C. 2007a. Recursion, language, and starlings. *Cognitive Science* 31, 697–704.

Corballis, Michael C. 2007b. On phrase structure and brain responses: A comment on Bahlmann, Gunter, and Friederici (2006). *Journal of Cognitive Neuroscience* 19, 1581–1583.

Elman, Jeffrey L. 1991. Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning* 7, 195–225.

Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71–99.

Epstein, Samuel David & Norbert Hornstein. 2004. The future of language (science). *Language* 81, 3–6.

Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã.

*Current Anthropology* 46, 621–646.

Fitch, W. Tecumseh & Marc D. Hauser 2004. Computational constraints on syntactic processing in a non-human primate. *Science* 303, 377–380.

Folia, Vasiliki, Julia Uddén, Christian Forkstam, Martin Ingvar, Peter Hagoort & Karl Magnus Petersson, K. M. 2008. Implicit learning and dyslexia. *Annals of the New York Academy of Sciences* 1145, 132–150.

Forkstam, Christian., Åsa Elwér, Martin Ingvar & Karl Magnus Petersson. 2008. Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research* 1221, 80–92.

Forkstam, Christian, Peter Hagoort, Guillén Fernández, Martin Ingvar & Karl Magnus Petersson 2006. Neural correlates of artificial syntactic structure classification. *Neuroimage* 32, 956–967.

Foss, Donald J. & Helen S. Cairns. 1970. Some effects of memory limitations upon sentence comprehension and recall. *Journal of Verbal Learning and Verbal Behavior* 9, 541–547.

Friederici, Angela D. 2004. Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences* 8, 245–247.

Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schubotz & Alfred Anwander. 2006. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences* 103, 2458–2463.

Gebhart, Andrea L., Elissa L. Newport & Richard N. Aslin. 2009. Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin and Review* 16, 486–490.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14, 225–248.

Gimenes, Manuel, François Rigalleau & Daniel Gaonac'h. 2009. The effect of noun phrase type on working memory saturation during sentence comprehension. *European Journal of Cognitive Psychology* 21, 980–1000.

Gomez, Rebecca L. 2002. Variability and detection of invariant structure. *Psychological Science* 13, 431–436.

Gomez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4, 178–186.

Greenfield, Patricia M. 1991. Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral Brain Sciences* 14, 531–595.

Greenfield, Patricia M., Karen Nelson & Elliot Saltzman. 1972. The development of rulebound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology* 3, 291–310.

Hagstrom, Paul & Jaemin Rhee. 1997. The Dependency Locality Theory in Korean. *Journal of Psycholinguistic Research* 26, 189–206.

Hakes, David T. & Helen S. Cairns. 1970. Sentence comprehension and relative pronouns. *Perception & Psychophysics* 8, 5–8.

Hakes, David T., Judith S. Evans, & Linda L. Brannon. 1976. Understanding sentences with relative clauses. *Memory & Cognition* 4, 283–290.

Hakes, David, T. & Donald J. Foss. 1970. Decision processes during sentence comprehension: Effects of surface structure reconsidered. *Perception & Psychophysics* 8, 413–416.

Hamilton, Helen W. & James Deese. 1971. Comprehensibility and subject–verb relations in complex sentences. *Journal of Verbal Learning &Verbal Behavior* 10, 163–170.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Hauser, Marc D., Elissa L. Newport & Richard N. Aslin. 2001. Segmentation of the speech stream in a non-human primate: Statistical learning in cottontop tamarins. *Cognition* 78, B53–B64.

Hochmann, Jean-Rémy, Mahan Azadpour & Jacques Mehler. 2008. Do humans really learn A*n*B*n* grammars from exemplars? *Cognitive Science* 32, 1021–1036.

Hoover, Michael L. 1992. Sentence processing strategies in Spanish and English. *Journal of Psycholinguistic Research* 21, 275–299.

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43, 365–392.

Heimbauer, Lisa A., Christopher M. Conway, Morten H. Christiansen, Michael J. Beran & Michael J. Owren. 2010. Grammar rule-based sequence learning by rhesus macaque (Macaca mulatta). Paper presented at the *33rd Meeting of the American Society of Primatologists*, Louisville, KY. [Abstract in *American Journal of Primatology* 72, 65].

Johnson-Pynn, Julie, Dorothy M. Fragaszy, Elizabeth M. Hirsch, Karen E. Brakke & Patricia M. Greenfield. 1999. Strategies used to combine seriated cups by chimpanzees (Pantroglodytes), bonobos (Pan paniscus), and capuchins (Cebus apella) *Journal of Comparative Psychology* 113, 137–148.

Knowlton, Barbara & Larry R. Squire. 1996. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 169–181.

Larkin, Willard & David Burns. 1977. Sentence comprehension and memory for embedded structure. *Memory & Cognition* 5, 17–22.

Lasnik, Howard. 2000. *Syntactic Structures Revisited: Contemporary Lectures on Classic Transformational Theory*. Cambridge, MA: MIT Press.

Levelt, Willem J. M. 1974. *Formal Grammars in Linguistics and Psycholinguistics: Vol. III, Psycholinguistic Applications*. The Hague: Mouton.

Liberman, Mark. 2004a. http://itre.cis.upenn.edu/~myl/languagelog/archives/000355.html.

Liberman, Mark. 2004b. http://itre.cis.upenn.edu/~myl/languagelog/archives/001399.html.

Lieberman, Matthew D., Grace Y. Chang, Joan Chiao, Susan Y. Bookheimer & Barbara J. Knowlton. 2004. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive*

*Neuroscience* 16, 427–438.

Lobina, David J. 2011. Recursion and the competence/performance distinction in AGL tasks. *Language and Cognitive Processes*, doi: 10.1080/01690965.2011.

Makuuchi, Michiru, Jörg Bahlmann, Alfred Anwander & Angela D. Friederici. 2009. Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences* 106, 8362–8367.

Marks, Lawrence E. 1968. Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior* 7, 965–967.

Miller, George A. 1962. Some psychological studies of grammar. *American Psychologist* 17, 748–762.

Miller, George A. & Stephen Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7, 292–303.

Minsky, Marvin L. 1967. *Computation: Finite and Infinite Machines.* Englewood Cliffs, NJ: Prentice-Hall.

Misyak, Jennifer B., Morten H. Christiansen & J. Bruce Tomblin. 2010a. On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, doi:10.3389/fpsyg.2010.00031.

Misyak, Jennifer B., Morten H. Christiansen & J. Bruce Tomblin. 2010b. Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science* 2, 138–153.

Müller, Jutta L., Jörg Bahlmann & Angela D. Friederici. 2010. Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science* 34, 338–349.

Newport, Elissa L. & Richard N. Aslin. 2004. Learning at a distance I: Statistical learning of nonadjacent dependencies. *Cognitive Psychology* 48, 127–162.

Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen & Richard N. Aslin. 2004. Learning at a distance II: Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49, 85–117.

Onnis, Luca, Padraic Monaghan, Korin Richmond & Nick Chater. 2005. Phonology impacts segmentation in speech processing. *Journal of Memory and Language* 53, 225–237.

Packard, Mark G. & Barbara J. Knowlton. 2002. Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience* 25, 563–593.

Pavlidou, Elpis V., Joanne M. Williams & Louise M. Kelly. 2009. Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of Dyslexia* 59, 55–77.

Perfors, Amy, Josh Tenenbaum, Edward Gibson & Terry Regier. 2010. How recursive is language? A Bayesian exploration. In Harry van der Hulst (ed.), *Recursion in Human Language*, 159–175. Berlin: Mouton de Gruyter.

Perruchet, Pierre & Sebastien Pacton. 2006. Implicit learning and statistical learning: Two approaches, one phenomenon. *Trends in Cognitive Sciences* 10, 233–238.

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

Peterfalvi, Jean M. & Francoise Locatelli. 1971. L'acceptabilité des phrases [The

acceptability of sentences]. *L' Année Psychologique* 71, 417–427.

Petersson, Karl Magnus. 2005. On the relevance of the neurobiological analogue of the finite state architecture. *Neurocomputing* 65–66, 825–832.

Petersson, Karl Magnus. 2008. On cognition, structured sequence processing and adaptive dynamical systems. *Proceedings of the American Institute of Physics*: *Mathematical and Statistical Physics Subseries* 1060, 195–200.

Petersson, Karl Magnus, Christian Forkstam & Martin Ingvar. 2004. Artificial syntactic violations activate Broca's region. *Cognitive Science* 28, 383–407.

Petersson, Karl Magnus, Vasiliki Folia & Peter Hagoort. 2010. What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, doi:10.1016/j.bandl.2010.08.003.

Pullum, Geoffrey K. & Barbara C. Scholz. 2010. Recursion and the infinitude claim. In Harry van der Hulst (ed.), *Recursion in Human Language*, 113–138. Berlin: Mouton de Gruyter.

Pylyshyn, Zenon W. 1973. The role of competence theories in cognitive psychology. *Journal of Psycholinguistic Research* 2, 21–50.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior* 6, 855–863.

Reber, Arthur S. 1989. Implicit learning and tacit knowledge. *Journal of Experimental Psychology General* 118, 219–235.

Reber, Arthur S. & Rhianon Allen. 1978. Analogic and abstraction strategies in synthetic grammar learning: A functional interpretation. *Cognition* 6, 189–221.

Reber, Paul J. & Larry R. Squire. 1999. Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience* 113, 235–242.

Rüsseler, Jascha, Ivonne Gerth & Thomas Münte. 2006. Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology* 28, 808–827.

Saffran, Jenny R. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44, 493–515.

Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. & Elissa L. Newport. 1999. Statistical learning of tone sequences by adults and infants. *Cognition* 70, 27–52.

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.

Siegelmann, Hava T. 1999. *Neural Networks and Analog Computation: Beyond the Turing Limit.* Basel: Birkhäuser.

Tallerman, Maggie, Frederick Newmeyer, Derek Bickerton, Denis Bouchard, Edith Kaan & Luigi Rizzi. 2009. What kinds of syntactic phenomena must biologists, neurobiologists, and computer scientists try to explain and replicate? In Derek Bickerton & Eörs Szathmáry (eds.), *Biological Foundations and Origin of Syntax*, 135–157. Cambridge, MA: MIT Press.

Thompson, Susan P. & Elissa L. Newport. 2007. Statistical learning of syntax: The

role of transitional probability. *Language Learning and Development* 3, 1–42.

Uddén, Julia, Susana Araujo, Christian Forkstam, Martin Ingvar, Peter Hagoort & Karl Magnus Petersson. 2009. A matter of time: Implicit acquisition of recursive sequence structures. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2444–2449.

Uddén, Julia, Vasiliki Folia, Christian Forkstam, Martin Ingvar, Guillén Fernández, Sebastiaan Overeem, Gijs Van Elswijk, Peter Hagoort & Karl Magnus Petersson. 2008. The inferior frontal cortex in artificial syntax processing: An rTMS study. *Brain Research* 1224, 69–78.

Uehara, Keiko. & Dianne Bradley. 1996. The effect of -ga sequences on processing Japanese multiply center-embedded sentences. In *11th Pacific-Asia Conference on Language, Information, and Computation*, 187–196. Seoul: Kyung Hee University.

Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92, 231–270.

van Heijningen, Caroline A. A., Jos de Visser, Willem Zuidema & Carel ten Cate 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Science* 106, 20538–20543.

Vasishth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25, 533–567.

de Vries, Meinou H., Andre R. C. Barth, Sandra Maiworm, Stefan Knecht, Pienie Zwitserlood & Agnes Flöel. 2010. Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *Journal of Cognitive Neuroscience* 22, 2427–2436.

de Vries, Meinou H., Padraic Monaghan, Stefan Knecht & Pienie Zwitserlood. 2008. Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition* 107, 763–774.

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

*Meinou H. de Vries*
*Vrije Universiteit*
*Department of Psychology*
  *and Education*
*Van der Boechorststraat 1*
*1081 BT Amsterdam*
*Netherlands*
*mh.de.vries@psy.vu.nl*

*Morten H. Christiansen*
*Cornell University*
*Department of Psychology*

*228 Uris Hall*
*Ithaca, NY 14853–7601*
*United States*
*christiansen@cornell.edu*

*Karl Magnus Petersson*
*Max Planck Institute for*
  *Psycholinguistics*

*Wundtlaan 1*
*6525 XD Nijmegen*
*The Netherlands*
*karl-magnus.petersson@mpi.nl*