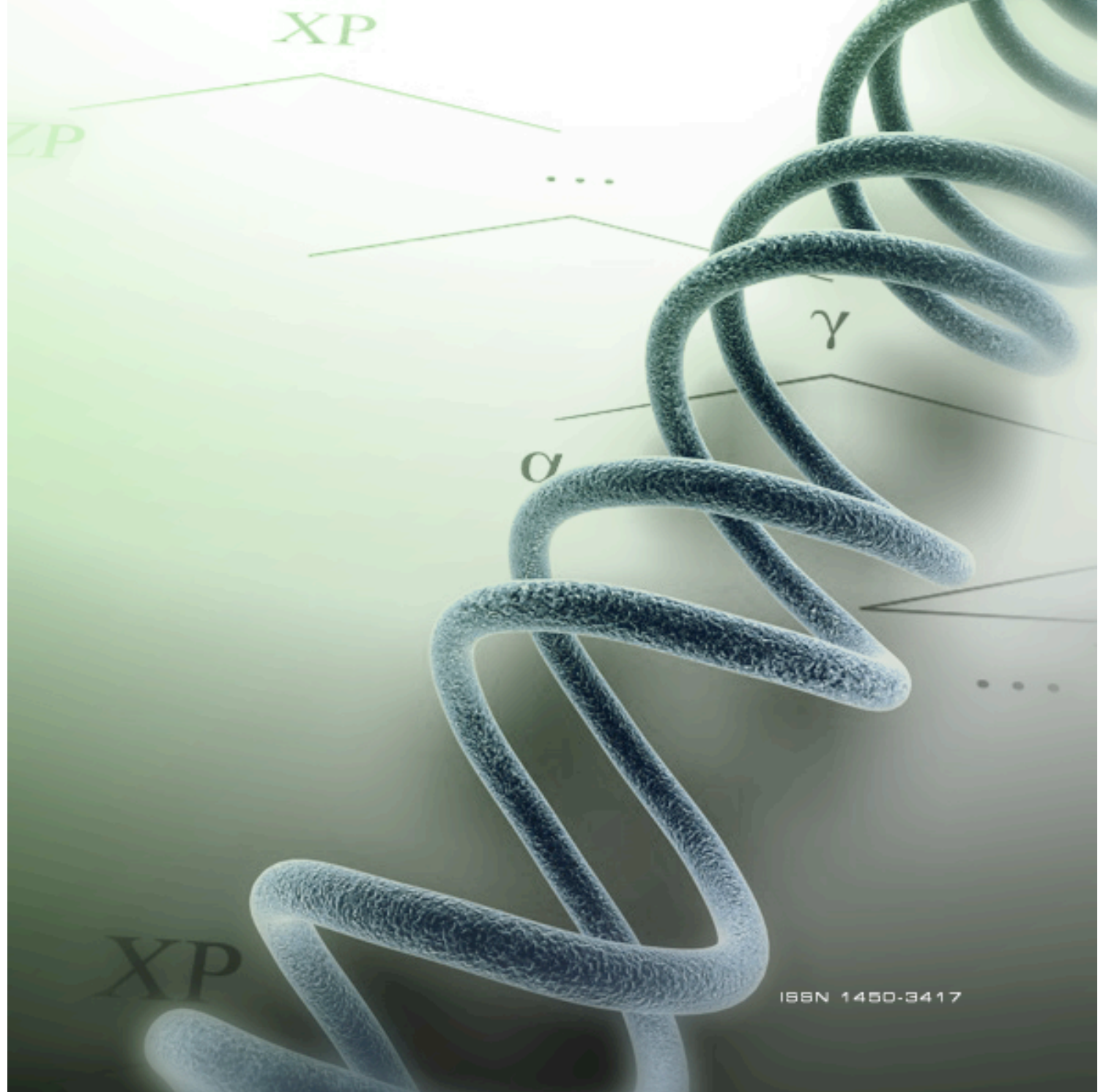




# biolinguistics

volume 5  
issue 1-2

winter-spring  
2011



ISSN 1450-3417



# Biolinguistic Perspectives on Recursion: Introduction to the Special Issue

Uli Sauerland & Andreas Trotzke

## 1. Introduction

Recursion is a central issue of the biolinguistic investigation of language. This special issue brings together seven new contributions on recursion from several different perspectives: theoretical syntax, neurolinguistics, language acquisition, genetics, and psycholinguistics. In this introduction, we first briefly characterize the background on recursion that the contributions share and that led to this special issue of *Biolinguistics*. Secondly, we situate the advances of the individual contributions in this issue against the background of studies on recursion.

The notion of recursion has played a significant role in the development of the field of linguistics and specifically of the generative approach. The concern that led to recursion is very old: Descartes (2003 [1637]: 38) hypothesized that the crucial difference between man and animal manifests itself most clearly in the fact that an animal “never [...] arranges its speech in various ways [...] in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do”. In a similar vein, and two centuries later, Wilhelm von Humboldt (1999 [1836]: 91) pointed out the human capacity to “make infinite employment of finite means” in language. Recursive rules provide one solution to the problem of accounting for an infinite number of possible sentences by means of a finite memory space. However, Descartes and Humboldt didn’t yet talk about recursion specifically, and infinity could be produced by means other than recursion. In the 1950s, Noam Chomsky developed formal language theory as a mathematically precise model of language and, using it, specified a precise role of recursion within formal models of language. In fact, recursion in one specific way proved to be essential to set apart the phrase structure models of language Chomsky proposed from the behaviorist models of language prevalent

---

Several of the papers in this volume originate from the workshop “[...] von endlichen Mitteln einen unendlichen Gebrauch machen” — Recursion as a Central Issue in Recent Linguistics’, held at the 2010 meeting of the German Linguistics Society (DGfS) at the Humboldt-University in Berlin. We thank the organizers of the DGfS-meeting for providing us with the venue and the participants of the workshop for many interesting discussions. In person, we thank Manfred Krifka for initiating the workshop and Rainer Ludwig and Mathias Schenner for co-organizing the workshop with us. We gratefully acknowledge financial support for the workshop and the completion of this volume from the German research foundation (DFG grant SA 925/1) and the European Union (project CHLaSC). We are grateful to all the scholars who shaped this volume as reviewers. Finally, we would like to thank Cedric Boeckx and Kleantes Grohmann for providing just the right platform to publish this special issue.



at the time. Chomsky, however, didn't use the term *recursion* for this notion, but defined the notion of *self-embedding* as follows:

- (1) A language  $L$  is *self-embedding* (s.e.) if it contains an  $A$  such that for some  $\phi, \psi$  ( $\phi \neq I \neq \psi$ ),  $A \Rightarrow \phi A \psi$ . (Chomsky 1959: 148)

The definition characterizes as self-embedding any language that contains a string  $A$  and allows the derivation from  $A$  of a string that properly contains  $A$ , that is,  $A$  is preceded and followed by two non-trivial strings. Over two papers, (Chomsky 1956, 1959) showed that the concept of self-embedding precisely sets apart context-free grammars from less complex models of grammar (specifically, finite state Markov process based models): All and only the languages produced by a context-free grammar that are self-embedding cannot be given an analysis using the less complex models.

Chomsky (1957) furthermore showed that English is self-embedding. In a nutshell, this demonstration consists of the observation that patterns such as (2) exist in English (slightly modified from Chomsky 1957: 22) and clearly satisfy the definition of self-embedding in (1).

- (2) a.  $S \Rightarrow$  If  $S$ , then it's true.  
b.  $S \Rightarrow$  Either  $S$  or not.

Finite state Markov chain models of language cannot capture the long-distance dependencies between *if* and *then* and *either* and *or*. Therefore, Chomsky established that behaviorist accounts of language were insufficient, whereas the phrase structure grammars Chomsky introduced could be sufficient. In this way, recursion was crucial for the development of phrase structure based approaches to language. However, subsequently recursion was not a major topic: Once phrase structures were established, recursion became part of the background. Chomsky and many other linguists proceeded to develop concrete phrase structure based grammars for specific languages.

Almost ten years ago, recursion became an active topic of research again due to work of Marc Hauser, Noam Chomsky, and Tecumseh Fitch. In an influential paper, Hauser *et al.* (2002) formulate a new hypothesis involving recursion. For this purpose, they differentiate between the broad and the narrow sense of faculty of language, drawing on the basic biolinguistic distinction between human traits that can be relegated to more general cognitive capacities, which, as Hauser *et al.* claim, are shared with other animals, and traits that are both human- and language-specific. They then hypothesize that only syntactic recursion belongs to the faculty of language in the narrow sense. With syntactic recursion, Hauser *et al.* seem to have in mind a general ability that underlies Chomsky's (1959) notion of self-embedding language in (1): a property of languages that distinguishes between phrase structure grammars and less powerful grammars (cf. Fitch 2010, Tomalin 2011, Luuk & Luuk 2011). Specifically, Hauser *et al.* write that "[n]atural languages go beyond purely local structure by including a capacity for recursive embedding of phrases within phrases" (p. 1577). Recursion is more general than self-embedding, though: The natural numbers, for example, also

rely on recursion, though possibly recursion within a finite-state grammar. Following much of the subsequent psychological literature (e.g. Gentner *et al.* 2006, Friederici *et al.* 2006), we assume that only a grammar that can account for self-embedding languages should be called *recursive*, and will use the term in this restricted sense in the following.

The hypothesis of Hauser *et al.* and their arguments captured the imagination of many researchers from different disciplines. Even though an enormous amount of progress has been made, many of the debates Hauser *et al.* triggered are still not resolved. This special issue takes up three major concerns that have developed since 2002. The first concern is how to test for recursion in experimental psychology: Since we cannot test humans on infinite sets of sentences and furthermore self-embedding in the sense of (1) is difficult for humans to process, how can recursion best be tested for? The second major concern is the role of recursion in linguistic theory: Is recursion an integral part of any syntactic structure building or is recursion better viewed as something on top of more basic structure building? Finally, the third major concern this issue addresses is the relation of recursion to the genetic and neural basis of language. Can recursion be separated from other parts of language in the genetic and neural domain?

## 2. Testing Recursion in Experimental Psychology

Testing recursion in experimental psychology crucially rests on the Artificial Grammar Learning (AGL) paradigm, which goes back to Reber (1967) and enables psychologists to isolate fundamental mechanisms involved in natural language syntax in sophisticated test designs. The first AGL study interpreted to be relevant for understanding syntactic recursion at the behavioral level was carried out by Fitch & Hauser (2004). They focused on the comparison between a self-embedding language and one that isn't. The two types of structures can be illustrated with the two English sentences in (3) and (4):

(3) [The man]<sub>A</sub> [the dog]<sub>A</sub> [bit]<sub>B</sub> [comes]<sub>B</sub>.

(4) [The man]<sub>A</sub> [comes]<sub>B</sub>; [the dog]<sub>A</sub> [bit the man]<sub>B</sub>.

As already mentioned, a finite-state grammar cannot account for center-embedded languages. Fitch & Hauser created stimuli that correspond to the artificial grammar  $A^nB^n$ , generating structures like (3) with  $n=2$ , and  $(AB)^n$ , yielding for  $n=2$  structures like (4). The actual stimuli were not created using English words, however, but classes of syllables. With these stimuli, Fitch & Hauser compared the parsing abilities of cotton-top tamarins (*Saguinus oedipus*) and humans regarding both grammar types. The result of their study was "that tamarins suffer from a specific and fundamental computational limitation on their ability to spontaneously recognize or remember hierarchically organized acoustic structures" (p. 380). That is, while tamarins were able to process structures generated by the  $(AB)^n$  grammar, they were not capable of mastering structures according to the  $A^nB^n$  formula. Accordingly, this experimental study supports the hypothesis that "the acquisition of hierarchical processing ability

may have represented a critical juncture in the evolution of the human language faculty" (p. 380) and thus may be of direct relevance to the hypothesis suggested by Hauser *et al.* (2002).

However, a finite set of experimental materials such as that of Fitch & Hauser couldn't in principle exclude all alternative explanations other than a recursive phrase structure grammar of the performance of the human subjects. For the materials of Fitch & Hauser, specifically, humans may have relied on counting rather than grammar building when processing the A<sup>n</sup>B<sup>n</sup> sequences or on additional cues specific to their stimuli. Research since has explored different experimental methods and formal grammars to more precisely pin down specific human abilities. One focus has been the fact that natural human language requires the ability to process sequences in which a consistent coupling of AB-pairs is involved (cf. Perruchet & Rey 2005, de Vries *et al.* 2008). To visualize, the more exact representation of our sentence (3), according to this objection, must be (5), where the pairing of particular As and Bs is marked by numbers:

(5) [The man]<sub>A1</sub> [the dog]<sub>A2</sub> [bit]<sub>B2</sub> [comes]<sub>B1</sub>.

The results for formal grammars eliciting such structures have been mixed. Given several methodological issues, it is now a central concern in this field of inquiry "that the relation between artificial language studies and natural language must be clarified" (Hauser *et al.* 2007: 127). As a consequence, experimental methods and insights have become more differentiated. Two contributions in this issue, one by de Vries *et al.* and one by Poletiek, advance this agenda.

**Meinou de Vries, Morten H. Christiansen, and Karl Magnus Petersson** argue that research focusing only on nested dependencies like (3) cannot provide us with a complete picture of where the boundaries of human language processing lie. They show that crossed dependencies, another type of non-adjacent dependencies, are easier to learn than nested dependencies, if the number of dependencies exceeds two. In light of this finding, they argue that the different complexity levels formulated in the Chomsky hierarchy (cf. Chomsky 1956) and used by studies such as Fitch & Hauser (2004) and Gentner *et al.* (2006) are less relevant. Instead, they propose a new complexity hierarchy, which is based on the assumption that syntactic complexity is determined by (i) the number of dependencies that need to be resolved and (ii) the specific ordering of these dependencies.

**Fenna Poletiek** argues that the so-called 'staged input effect' is relevant for learning an artificial grammar with center-embedded structures. Referring to studies that have shown that artificial grammars with center-embedded structures are difficult to learn by induction, Poletiek claims that participants do better if the input used to train human subjects is presented in an incremental organization, starting with the least complex and ending with the most complex exemplars. Crucially, this staged input effect is argued to be helpful for hierarchical structures only and shows no effect for learning a finite state grammar. Like de Vries *et al.*, Poletiek advances the perspective that competence models for language complexity like the Chomsky hierarchy should care about performance factors such as learnability.

### 3. Locating Recursion in Linguistic Theory

Syntactic theories describe generalizations at various levels of the typology of phrases. At the least specific level of the phrasal typology, any word or phrase belongs to the same type. All that can be said for this type of phrase is that two phrases can form a complex syntactic constituent, as is captured by the operation *Merge* in minimalist syntax. Many syntactic generalizations, however, make reference to a more articulated typology of phrases. For example, a generalization across several languages is that single words (heads) are distinguished from complete phrases for word order phenomena: Languages can therefore be described as either head-initial (e.g. English) or head-final (e.g. Japanese). This supports a typology of phrases that distinguishes between heads (single words) and maximal projections (complete phrases). A second distinction between types of phrases is important within minimalist syntax: the distinction between phases and non-phases. Phases are special phrases that are distinct from other phrases by their intonational and semantic properties.

The debate over the appropriate typology of phrases does not directly relate to the formal notion of self-embedding as characterized in (1), since the formal notion applies to languages — not to concrete grammars. However, there exists also a natural notion of recursion that applies to specific structures in a phrase structure based syntax: A structure is recursive if there is a phrase of type X that contains as a proper part another phrase of type X. Recursion of a concrete phrase structure grammar is evidently a different notion from self-embedding as defined in (1). The link between the notions is the following: If a language is self-embedding, any phrase structure grammar must be such that some strings are analyzed as having a recursive structure. But, importantly, the notion of self-embedding never predicts which specific strings must receive a recursive analysis, nor is any language that can be analyzed with phrase-structure grammar that allows some recursive structures necessarily self-embedding in the formal sense.

The notion of recursive structure is nevertheless an important one, especially since for natural language other sources of evidence (for example, intonational and semantic evidence) are available to determine the phrase structure of a specific sentence. Which structures are recursive, however, is closely tied the typology of phrases. As already mentioned, current minimalist syntax assumes an abstract operation *Merge* as the only phrase structure rule. *Merge* always structures exactly two items into one phrase. Therefore, any sentence consisting of three or more words must involve a recursive operation of phrase structure building on this view, as Nevins *et al.* (2009) point out. But if the typology of phrases assumed is richer, a smaller set of structures are analyzed as recursive. For example, the traditional phrase structure rules  $S \Rightarrow NP V$  and  $NP \Rightarrow D N$  allow the analysis of the sequence D N V without recursion. Moreover, some scholars committed to more ‘strong’ derivational approaches like phase theory or other models implying multiple points of Spell-Out have recently argued that the narrow structure-building operations of grammar are not recursive at all and that recursion might better be described as an interface phenomenon (cf. Arsenijević & Hinzen 2010, Surányi 2010).

**Jan-Wouter Zwart** assumes a minimalist background and claims that syntactic recursion should not be defined in terms of embedding, but in terms of derivation layering. Comparing iterative and recursive procedures to build phrase structures, he argues that one cannot decide that a language is recursive by simply looking at its structures. Instead, one has to investigate the structure building procedure itself. After showing that embedding structures can also be generated without recursion, Zwart defines recursion in language as the interaction between derivation layers. He then applies these concepts to the analysis of the Amazonian language Pirahã. Everett (2005) claims that Pirahã does not exhibit recursion which has led to intense discussion on the relevance of this finding with respect to the status of recursion as a linguistic universal. According to Zwart's approach, both complex subjects and structured lexical items imply recursion. Zwart then demonstrates, using uncontested data, that both complex subjects and structured lexical items are attested in Pirahã, and thus, he argues that the grammar of Pirahã allows for recursive structures.

**Tom Roeper** starts with the assumption that if variation is attested regarding what particular forms of recursion natural languages allow, then an acquisition challenge exists. In the light of acquisition evidence from adjectives, possessives, verbal compounds, and sentence complements, he outlines an acquisition path for specific forms of recursion. In particular, he distinguishes three mechanisms to build recursive structures: direct recursion, indirect recursion, and Generalized Transformations (GTs), as realized in an adaptation of Tree Adjoining Grammar. Since children first analyze adjacent identical structures as direct recursion with a conjunctive reading, Roeper argues that direct recursion is the acquisition default and can thus be viewed as the first stage in the acquisition of recursive structures. Assuming that children must 'experience' specific forms of recursion in order to allow them in their language, he goes on to discuss several evidences that may help account for the path of how to acquire the more complex forms of both indirect recursion and recursion in the form of GTs.

#### 4. Localizing Recursion in Cognitive Neuroscience and Genetics

The testing of human and non-human subjects regarding their capacities to process artificial language grammars inspired neuropsychological studies that ask to what extent the core computational faculty of processing hierarchical embedded structure can be segregated from other brain functions. Let us briefly look at this field of research.

Friederici *et al.* (2006) build on the findings of Fitch & Hauser (2004) and hence assume that humans differ from non-human primates in their capacity to master sequences that are generated by the  $A^nB^n$  grammar. In their study, they ask, broadly speaking, whether the differences of processing the two grammars used by Fitch & Hauser are reflected in the human brain. To explore this question, they test human subjects by visually presenting sequences of consonant-vowel syllables that were modeled to represent the different grammar types. After having used these stimuli and after having applied several sophisticated testing procedures, they indeed conclude that there are differences



in processing in the brain. In particular, Friederici *et al.* report that processing of local transitions within a finite-state grammar is subserved by the left frontal operculum, whereas a specific section of Broca's area holds responsible for the computation of hierarchical dependencies involved in syntactic recursion within a phrase structure grammar. However, like in experimental psychology, the testing methods have been refined in subsequent studies, which is the starting point for three contributions in the present volume.

**Angela D. Friederici, Jörg Bahlmann, Roland Friedrich, and Michiru Makuuchi** report such refinements by reviewing recent neuroimaging experiments that evaluate the neural basis of processing embedded structures, which, as they argue, allows for conclusions regarding the localization of processing recursion in the brain. Based on numerous studies, they conclude that a special region of Broca's area, left Brodmann area 44, is the neural correlate of computing linguistic recursion. They segregated this correlate from activation of Broca's area due to working memory, from activation due to the processing of visual-event sequences, and from areas involved in processing hierarchically structured mathematical formulae. Friederici *et al.*'s cross-study review thus suggests two different computational systems in the lateral prefrontal cortex dealing with hierarchical structures, one which is domain-general and is active when processing complex hierarchies in non-language domains, and one which is domain-specific and deals with recursive language or language-like hierarchies.

**Vasiliki Folia, Christian Forkstam, Martin Ingvar, Peter Hagoort, and Karl Magnus Petersson** compare the brain networks engaged in processing grammaticality judgments and in processing preference judgments in an artificial grammar learning experiment. Their results show that preference and grammaticality classification engage virtually identical brain regions. That is, the subjects also engage brain regions central to natural syntax processing when they are not explicitly instructed or receive any information concerning the existence of a grammatical rule system that underlies the presented stimuli. In addition, Folia *et al.* present some initial efforts to understand the genetic basis of the capacity for artificial syntax acquisition by exploring the potential role of the CNT-NAP2 gene, which is controlled by the FOXP2 transcription factor and whose expression is enriched in frontal brain regions in humans.

**Eleonora Russo and Alessandro Treves** ask what evolutionary changes have occurred in the human neocortex that allow for the crucial feature of recursion in human language. After having reviewed salient features of cortical organization, they discuss recent work that shows that the human cortex has more neurons (in absolute number) than any other mammal, or, more specifically, that the number of spines present on the dendrites of pyramidal cells are significantly higher in the human cortex than in any other species. They then argue that these quantitative differences can produce qualitative changes in the functionality of a neural network. Discussing the phenomenon of latching dynamics, that is, the 'hopping' of the network from one attractor state to another, they refer to their previous analyses of this phenomenon and sketch the boundary between finite and infinite latching. In particular, they claim that a network latches indefinitely when the memory load is above a certain critical value. By assuming that latching is a property that emerges when crossing

certain threshold, they suggest the evolutionary scenario that recursion in human language has evolved due to a slowly evolving quantitative increase in the connectivity of the cortex that has suddenly crossed a critical threshold. Accordingly, syntactic recursion may have emerged in a manner entirely unrelated to the appearance of a novel piece in the neural circuitry, that is, without altering the intrinsic make-up of the network.

In addition to the seven research papers, a book review by **David J. Lobina** is included in this volume. Lobina reviews the recently published volume *Recursion and Human Language*, edited by Harry van der Hulst. We feel the review should be read in addition to the editorial to elaborate further the background for the research presented in this volume.

## References

- Arsenijević, Boban & Wolfram Hinzen. 2010. Recursion as a human universal and as a primitive. *Biolinguistics* 4, 165–173.
- Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions of Information Theory IT-2* 3, 113–124.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1959. On certain formal properties of grammars. *Information and Control* 2, 137–167.
- Descartes, René. 2003 [1637]. Discourse on method. In Elizabeth S. Haldane & George R. Thomson (trans.), *Discourse on Method and Meditations*, 1–52. Mineola, NY: Dover.
- 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.
- Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46, 621–634.
- Fitch, W. Tecumseh. 2010. Three meanings of ‘recursion’: Key distinctions for biolinguistics. In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*, 73–90. Cambridge: Cambridge University Press.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- 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 of the United States of America* 103, 2458–2463.
- Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- 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., David Barner & Tim O'Donnell. 2007. Evolutionary linguistics: A new look at an old landscape. *Language Learning & Development* 3, 101–132.
- von Humboldt, Wilhelm. 1836/1999. The diversity of human language-structure and its influence on the mental development of mankind. In Peter Heath (trans.), *Wilhelm von Humboldt: On Language*, 1–287. Cambridge: Cambridge University Press.
- Luuk, Erkki & Hendrik Luuk. 2011. The redundancy of recursion and infinity for natural language. *Cognitive Processing* 12, 1–11.
- Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85, 355–404.
- Perruchet, Pierre & Arnaud Rey. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review* 12, 307–313.
- Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior* 6, 855–863.
- Surányi, Balázs. 2010. Toward a strongly derivational syntax. In Michael T. Putnam (ed.), *Exploring Crash-Proof Grammars*, 167–212. Amsterdam: John Benjamins.
- Tomalin, Marcus. 2011. Syntactic structures and recursive devices: A legacy of imprecision. *Journal of Logic, Language and Information*, doi:10.1007/s10849-011-9141-1.

Uli Sauerland  
Zentrum für Allgemeine Sprachwissenschaft (ZAS)

Schützenstraße 18  
10117 Berlin  
Germany  
[uli@alum.mit.edu](mailto:uli@alum.mit.edu)

Andreas Trotzke  
Universität Konstanz  
Fachbereich Sprachwissenschaft  
Universitätsstraße 10  
78457 Konstanz  
Germany  
[andreas.trotzke@uni-konstanz.de](mailto:andreas.trotzke@uni-konstanz.de)

# 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)

---

This work was supported by the Netherlands Organisation for Scientific Research (NWO; project number 446-08-014), the Max Planck Institute for Psycholinguistics, the Donders Institute for Brain, Cognition and Behaviour, the Stockholm Brain Institute, the Fundação para a Ciência e Tecnologia (PTDC/PSI-PCO/110734/2009; IBB/CBME, LA, FEDER/POCI 2010), Vetenskapsrådet, the Hedlunds Stiftelse, and the Stockholm County Council (ALF, FoUU). We are very grateful to Vasiliki Folia, Julia Udden, and two anonymous reviewers for their helpful comments.



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.

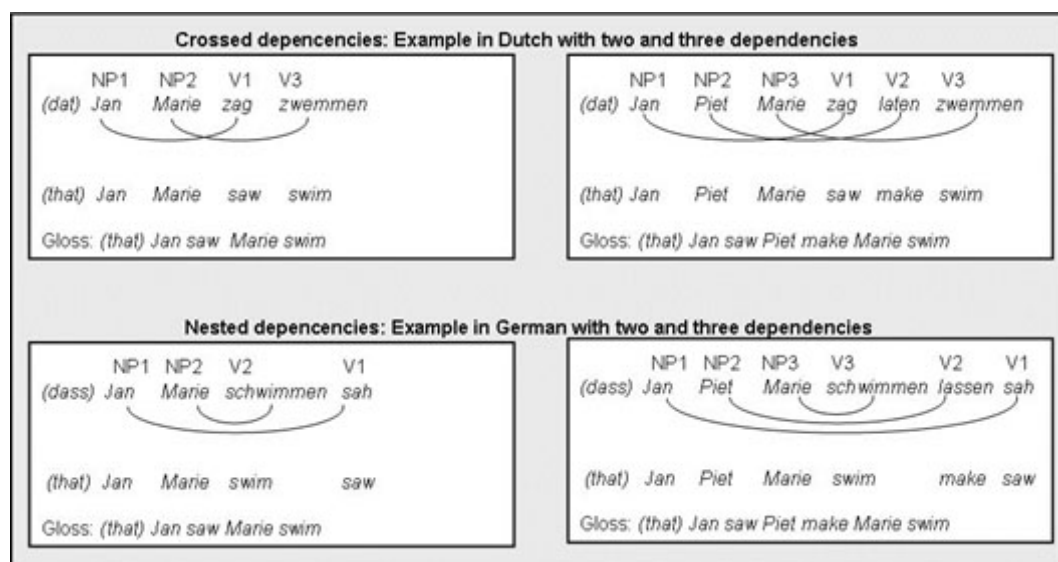


Figure 1: Different ways of expressing non-adjacent dependencies in German and Dutch

### 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.,  $x \rightarrow AxB$  and  $x \rightarrow \emptyset$  recursively defines  $\emptyset$ , AB, AAB, AAAB, AAAAB, 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 fledged 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, Blaugbergs & 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).

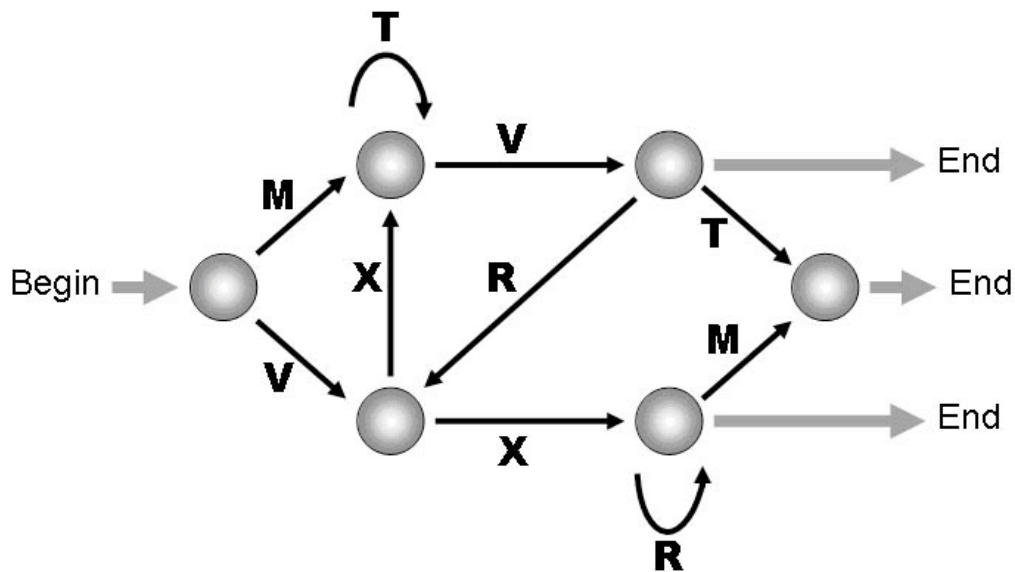


Figure 2. A finite-state grammar used to generate stimuli for artificial grammar learning tests (cf. Reber & Allen 1978).

## 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 remembering the order in which words occur in sentences, generalization of gram-

mathematical 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 extent, 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).



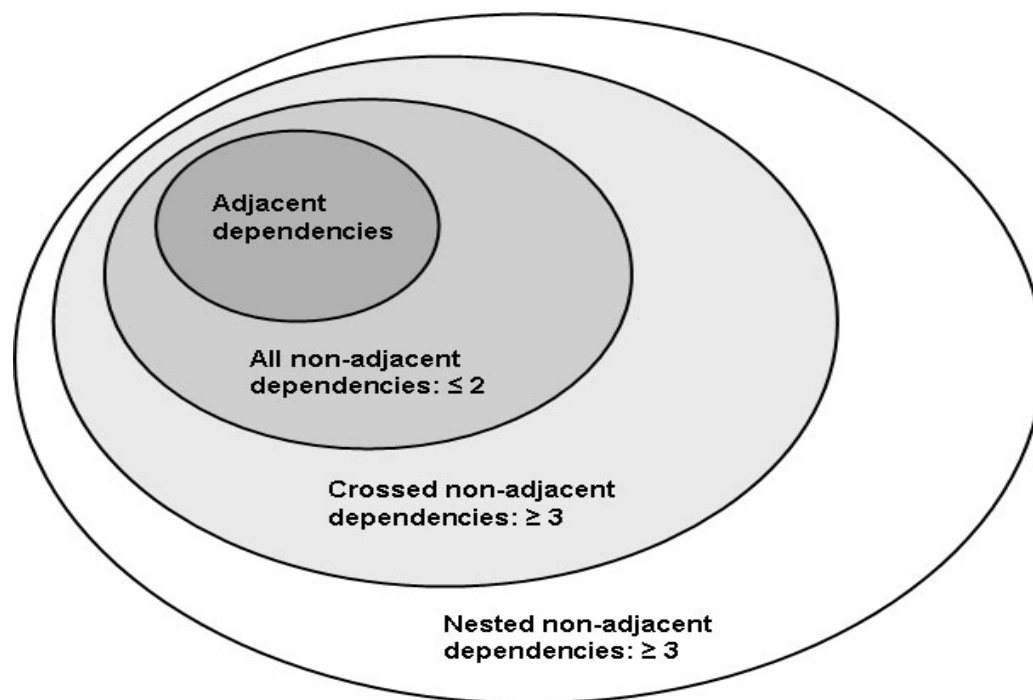


Figure 3: A schematic overview of our suggested levels of processing complexity. Note that each level of this hierarchy denotes a decisive factor that adds to processing complexity. Thus, it is not the case that three nested dependencies are harder to process than, say, six crossed dependencies. Instead, 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 & MacDonald (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.
- Blaubecks, 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 cotton-top tamarins. *Cognition* 78, B53–B64.
- Hochmann, Jean-Rémy, Mahan Azadpour & Jacques Mehler. 2008. Do humans really learn  $AnBn$  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 (*Pan troglodytes*), 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/language-log/archives/000355.html>.
- Liberman, Mark. 2004b. <http://itre.cis.upenn.edu/~myl/language-log/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 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](mailto: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](mailto:christiansen@cornell.edu)

Karl Magnus Petersson  
 Max Planck Institute for  
 Psycholinguistics  
 Wundtlaan 1  
 6525 XD Nijmegen  
 The Netherlands  
[karl-magnus.petersson@mpi.nl](mailto:karl-magnus.petersson@mpi.nl)

# What in the World Makes Recursion so Easy to Learn? A Statistical Account of the Staged Input Effect on Learning a Center-Embedded Structure in Artificial Grammar Learning (AGL)

Fenna H. Poletiek

In an artificial grammar learning study, Lai & Poletiek (2011) found that human participants could learn a center-embedded recursive grammar only if the input during training was presented in a staged fashion. Previous studies on artificial grammar learning, with randomly ordered input, failed to demonstrate learning of such a center-embedded structure. In the account proposed here, the staged input effect is explained by a fine-tuned match between the statistical characteristics of the incrementally organized input and the development of human cognitive learning over time, from low level, linear associative, to hierarchical processing of long distance dependencies. Interestingly, staged input seems to be effective only for learning hierarchical structures, and unhelpful for learning linear grammars.

*Keywords:* artificial grammar learning; linguistic environment; recursion; staged input; statistical learning

## 1. Recursion Learning in the Artificial Grammar Learning Paradigm

Language acquisition is one of the most complex tasks imaginable. Young learners, from infancy on, are faced with a noisy, degraded, and small set of streams of sounds — linguistic stimuli — from which grammatical principles have to be induced. Though generalization from the stimuli is needed to learn the grammar, it is bound to complex constraints: It should not go too far, and not be simple and linear. It is one of the most persistent mysteries in cognitive science how humans achieve this goal. How does this learning proceed?

Infants have been observed to induce simple linear statistical structure in a stream of sounds (Saffran 2003). Older children, however, induce highly complex non-linear rules from what they hear. For example, children never erroneously transform a sentence like ‘The man who was here yesterday is Sam’ into the corresponding question ‘Was the man who here yesterday is Sam?’ by simply moving the first encountered subordinate clause verb *was* to the front rather than the main verb *is* (Gómez & Gerken 2000). Moreover, in a statement like ‘The man



the dog bites shouts', the first encountered noun (subject) is associated with the last verb rather than with the first encountered next verb, revealing the application of a hierarchical principle.

The non-linear process required for natural language seems hard to explain with statistical learning mechanisms. Recently (and less recently; see Gold 1967), it has been proposed that this type of hierarchical structures is unique for human language and therefore is a crucial characteristic of the human language faculty (Hauser *et al.* 2002, Fitch & Hauser 2004; see also Corballis 2007). Very little is known, however, about how these structures are actually learned and used, and to what extent general statistical learning mechanisms and the learner's environment factor into acquiring hierarchical structures like center embedding.

The purpose of the present paper is to propose an explanation for a recently found facilitation effect of the organization of the linguistic input on learning a center-embedded structure (Lai & Poletiek 2011). The effect is accounted for in terms of the match between the statistical characteristics of the input and the developmental pattern of the learning process. I propose that the changing organization of the linguistic environment over time narrowly matches the synchronic development of cognitive learning mechanisms, binding formal language complexity to learnability.

Because of the extremely high complexity of natural grammars, little information about fundamental mechanisms involved in natural grammar learning can be derived directly from the features of language. Therefore, a growing body of research on grammar learning uses artificial grammars for both simulation studies and empirical experimentation. A now classical paradigm is the Artificial Grammar Learning (AGL) procedure (Reber 1969, 1993). Reber (1969) exposed human participants to exemplars of a simple finite state grammar (see Figure 1a below) with a few 'words' (mostly letters). Next, participants are given a test task in which new strings are presented, half of which are grammatical and half are not. Participants give grammaticality judgments for each test string. Typically, participants perform significantly above chance level, indicating that the structure was induced during training and applied during the test phase, at least to some extent.

The artificial grammar learning paradigm can be used to perform a laboratory test of possible effects of environmental characteristics on the learnability of sequential structures, by simulating these characteristics in the experimental task and comparing learning behavior under experimental conditions with a matched control condition in which the investigated characteristics are not implemented.

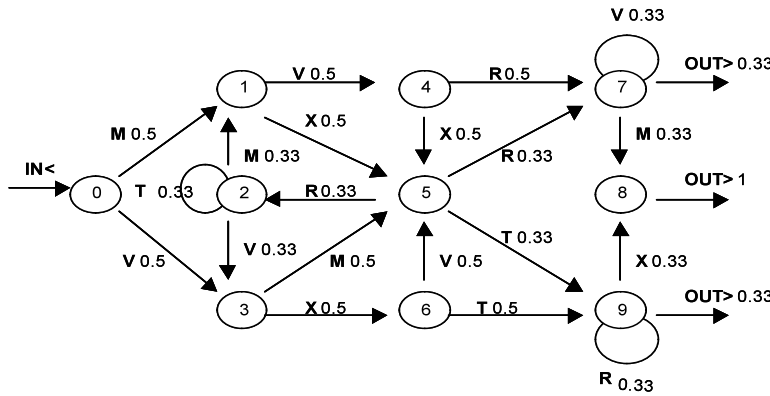
## **2. Staged Input Facilitates Hierarchical Processing**

Artificial grammars with a center-embedded rule have been shown to be extremely hard to learn by induction. Participants failed to show any knowledge of the hierarchical center-embedded structure after exposure to a randomly ordered set of exemplars (de Vries *et al.* 2008). In Lai & Poletiek's (2011) study, Reber's (1969) procedure was slightly adapted. Rather than presenting one learning and one test phase, the task was divided in twelve blocks each with a set of twelve

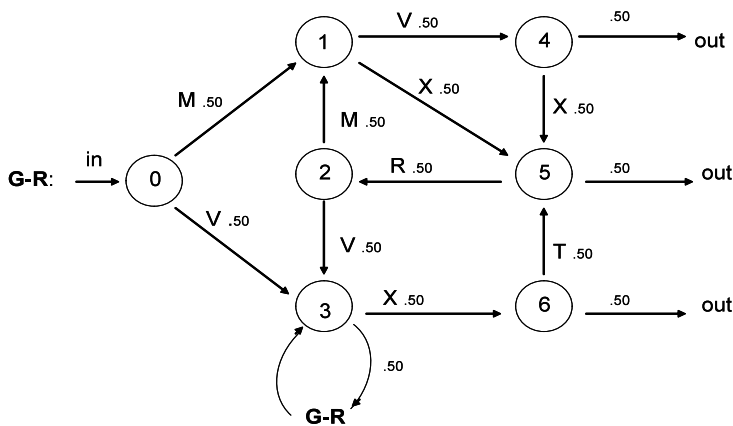
training strings followed by twelve test strings. This procedure allowed us to measure the development over time of grammaticality judgments performance as exposure increases.

In contrast to what de Vries *et al.* (2008) found, our participants performed well — but only if the input exemplars with which they were trained were presented in an incremental fashion, starting with the shortest and least complex exemplars without embeddings and ending with exemplars with multiple levels of embedding. This result suggests that the time course of exposure to hierarchically increasingly complex stimuli, allows cognitive learning of the hierarchical system. A statistical analysis of the input presented incrementally may provide an account of this staged input effect. Moreover, as I will argue below, this analysis also provides an explanation of why hierarchical structures rather than simple linear finite-state structures benefit from an incrementally organized input.

Consider a finite-state structure (G-FS) with five elements (letters M, V, R, X, T) and a hierarchical recursive center-embedded structure (G-R), with the same five elements:



a) Finite-state system G-FS



b) Center-embedded system G-R

Figure 1: Schematic probabilistic Markov-representation of a) artificial finite-state grammar (G-FS) and b) center-embedded recursive structure (G-R).



Both systems generate strings of elements (e.g. MXTRR for G-FS and VMXRMVX for G-R). For both systems, the probability of each unique element the system generates ( $p(\text{exemplar} | G)$ ) can be calculated (van der Mude & Walker 1978, Charniak 1994, Poletiek & Wolters 2009). The sum of the probabilities of all unique strings generated by a system (i.e. the ‘full output’) is, or approximates, one (Poletiek & van Schijndel 2009). The probability distributions of the unique strings generated by both grammars differ. Indeed, it can be shown that the probability distribution of the exemplars generated by G-FS is more ‘even’ than the probability distribution of the strings generated by the recursive structure G-R. In the recursive system, the strings without any center-embedded clause are much more probable than strings with embeddings. Moreover, as the number of levels of embedding increases, the probability of production by the system drops quickly to approximate zero.

If the exemplars of both grammars are ordered in a staged fashion, according to their decreasing probabilities to be produced by G-R and G-FS, then let us assume that learning of the grammars at any point in time may be represented by the sum of the probabilities of the exemplars a learner has been exposed to up to that point in time (Lai & Poletiek 2011). For example, if this sum ( $\Sigma(p(\text{exemplar} | G))$ ) is .50 after exposure to  $n$  exemplars presented in a growing fashion (staged input), the learner has been exposed and allowed to learn ‘half’ of the system. In Figure 2, the evolution over time of  $\Sigma(p(G | \text{string}))$  is displayed for both G-FS and G-R, for an input presented over time according to decreasing probabilities of the exemplars. Assuming that the cumulative probabilities of the gradually increasing set of input stimuli reflect the proportion of the full language (i.e. 100% of the stimuli it generates) at each stage of exposure, Figure 2 displays the evolution of this cumulative value over time for a Finite State Grammar and for a Recursive Grammar.

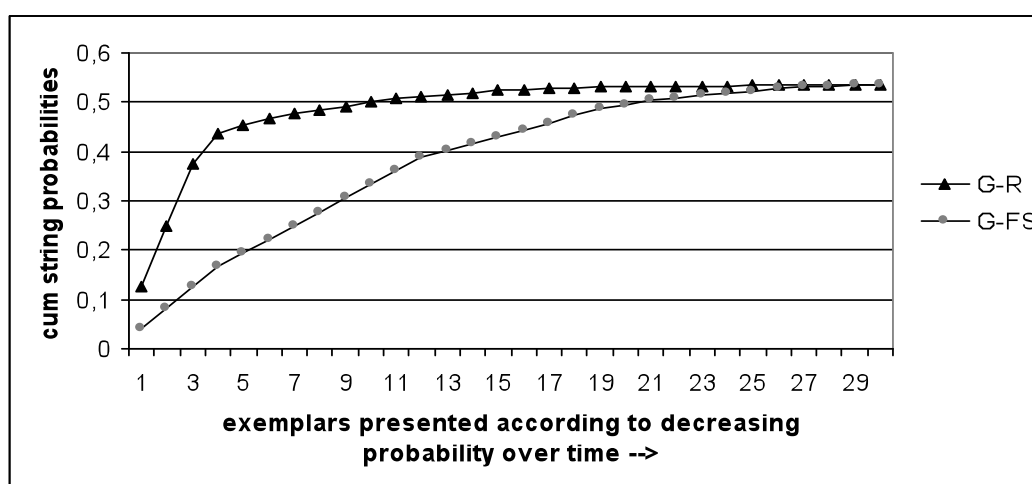


Figure 2: Cumulative exemplar probabilities for exemplars of the grammars G-FS and G-R ranked according to decreasing probabilities.

Consider two learners. After exposure to the 30 most probable exemplars of G-FS learner A has learned 50% of that grammar (Figure 2). Likewise, learner B, after having been exposed to the 30 most probable exemplars of G-R would also have learned 50% of G-R. Thus, the two learners have seen an equal number of exemplars of their ‘own’ grammar, covering an equal part of the grammar generating them. The difference between the two learning situations, however, is in the shapes of the curves.

As Figure 2 shows, in the earliest stage of exposure (e.g., after having seen five exemplars), the proportion (Poletiek & van Schijndel 2009) of the recursive system G-R, covered by the exemplars, exceeds by far the proportion of G-FS after exposure to five exemplars of G-FS. Assuming that the cumulative probabilities curve (y-values) reflects how much of the underlying system has been learned at each point of exposure (x-values), the lines might be considered to model learning curves of the two learners after a given amount of exposure. Figure 2 then reveals that presenting the input in an incremental fashion strongly boosts the learning curve of the *recursive* language in the *early stage* of learning (see also Elman 1993), according to this simulation.

This facilitation effect of staged input for recursive grammars was verified by Lai & Poletiek (2011) in their AGL study. Interestingly, the incremental presentation of the input does not help much for learning a linear finite-state grammar. As can be seen in Figure 2, the non-recursive grammar produces a more linear learning curve, implying a weaker effect of the organization of the input over time for non-recursive linear systems.

### 3. Artificial and Natural Grammar Learning

Translating this analysis to natural grammar learning requires a mapping of the artificial situation displayed in Figure 2 onto a natural developing language learner and natural linguistic input. A number of arguments can be advanced for the correspondence between the artificial data analysis and natural language learning. First, cognitive learning generally is time-course sensitive (Pine 1994). Not only language is acquired most effectively in the first years of life — also most skills and cognitive functions are learned best when we are young. Moreover, learning mechanisms are mainly statistical and associative during early childhood (Saffran 2003), becoming increasingly sophisticated and covering long-distance dependencies in later stages of learning. Accordingly, during the early stage of life, when the child is exposed to basic and short exemplars of the structure, cognitive processing is simple, linear, and associative, providing important information about the basic rules of the structure. Using this basic knowledge as a stepping stone, the learner’s growing cognitive capacity can process increasingly complex non-linear operations which allow the detection of recursive patterns.

Second, the staged environment assumed in the present artificial world may be argued to represent the linguistic natural environment of a young learner. Indeed, as studies into child-directed speech suggest (Pine 1994), linguistic utterances children are exposed to are simpler, shorter, and contain more frequent constructions than adult language. Only later on is the system to be in-

duced by the natural learner hierarchical and recursive — not linear.

In sum, the present theoretical explanation of the beneficial effect of a staged linguistic input on grammar induction, derived from artificial grammar study, suggests a well-tuned fit between the organization of the linguistic environment and the development of learning abilities. In addition, the model can explain why this facilitation occurs specifically for recursive grammars, and not for linear ones. Generally, as shown in this analysis, artificial grammar studies and statistical models of the effects they reveal are useful tools to our understanding of fundamental processes underlying natural grammar learning.

## References

- Charniak, Eugene. 1994. *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Corballis, Michael C. 2007. Recursion, language, and starlings. *Cognitive Science* 31, 697–704.
- Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71–99.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a non-human primate. *Science* 303, 377–380.
- Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10, 447–474.
- Gómez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4, 178–186.
- 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.
- Lai, Jun & Fenna H. Poletiek. 2011. The Impact of Adjacent-Dependencies and Staged-Input on the Learnability of Center-Embedded Hierarchical Structures. *Cognition*, 118(2), 265-273
- Pine, Julian M. 1994. The language of primary caregivers. In Clare Gallaway & Brian J. Richards (eds.), *Input and Interaction in Language Acquisition*, 15–37. Cambridge: Cambridge University Press.
- Poletiek, Fenna H. & Tessa J. P. van Schijndel. 2009. Stimulus set size and grammar coverage in artificial grammar learning. *Psychonomic Bulletin & Review* 16, 1058–1064.
- Poletiek, Fenna H. & Gezinus Wolters. 2009. What is learned about fragments in artificial grammar learning: A transitional probabilities approach. *Quarterly Journal of Experimental Psychology* 62, 868–876.
- Reber, Arthur S. 1969. Transfer of syntactic structures in synthetic languages. *Journal of Experimental Psychology* 81, 115–119.
- Reber, Arthur S. 1993. *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.
- Saffran, Jenny R. 2003. Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12, 110–114.

- van der Mude, Antony & Adrian Walker. 1978. On the inference of stochastic regular grammars. *Information and Control* 38, 310–329.
- 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.

*Fenna H. Poletiek*  
*Leiden University*  
*Cognitive Psychology Department*  
*P.O. Box 9555*  
*2300 RB Leiden*  
*The Netherlands*  
[poletiek@fsw.leidenuniv.nl](mailto:poletiek@fsw.leidenuniv.nl)

# Recursion in Language: A Layered-Derivation Approach

Jan-Wouter Zwart

This paper argues that recursion in language is to be understood not in terms of embedding, but in terms of derivational layering. A construction is recursive if part of its input is the output of a separate derivational layer. Complex clauses may be derived recursively in this sense, but also iteratively, suggesting that standard arguments for or against recursion in language are misdirected. More generally, we cannot tell that a grammar is recursive by simply looking at its output; we have to know about the generative procedure. Using the new definition of recursion in terms of derivational layering, we once again inspect the recorded data of Pirahã, arguing that there is reason to believe that the grammar of Pirahã is recursive after all.

*Keywords:* cyclicity; derivation; layered derivation; Pirahã; recursion

## 1. Spotting Recursion

A procedure is recursive if part of it involves running the entire procedure anew. The output of such a procedure is also called recursive. For example, the procedure that draws the famous Droste can picture in (1) below at some point involves calling the very same procedure, in order to draw that can *in* the picture.

However, one cannot tell that an object is recursive by simply looking at it: One has to know by what procedure it was derived. For example, the picture in (1) could have been generated in a non-recursive way (and it probably was), for instance by assigning a color to each pixel.

In another example, consider the situation in (2), from the comic strip ‘Calvin and Hobbes’, where the boy Calvin finds himself surrounded by five duplicates of

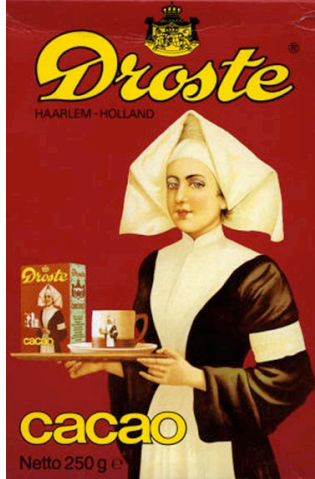
---

This paper was presented at the *TIN-dag* (Utrecht, 6 February 2010). I would like to thank the audience for their comments as well as Cedric Boeckx, Dan Everett, and two anonymous reviewers. The research reported on here is part of the project ‘Dependency in Universal Grammar’ sponsored by the Netherlands Organisation for Scientific Research (360-70-200), which is gratefully acknowledged.

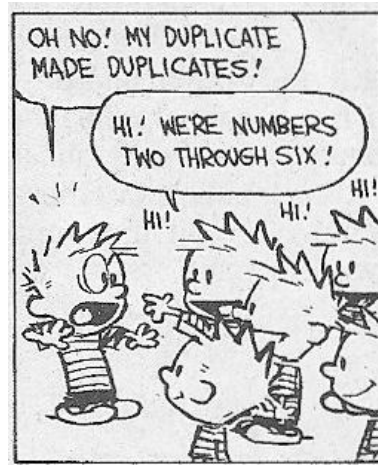


himself. As Calvin exclaims, the procedure bringing this situation about was recursive: Calvin made a duplicate of himself, and the duplicate proceeded to make a duplicate of *himself*, etc. However, Calvin could have created five duplicates by simply repeating the duplication process five times, yielding an iterative rather than a recursive procedure.

(1)



(2)



In language, it is commonly taken for granted that structures involving some kind of embedding are recursive. Typical examples are in (3):

- (3) a. John thinks that Mary said that it was raining.  
 b. The height of the letters on the cover of reports from the government.

The structures in (3) are recursive if they are generated by a recursive procedure, such as the rewrite rules of pre-minimalist generative grammar. The procedure generating (3a) may be paraphrased as in (4), showing recursion in that to interpret 'clause' in (4b), one must run (4a) again.

- (4) a. clause = subject + predicate  
 b. predicate = verb + clause

Similarly with the procedure in (5) generating (3b):

- (5) a. determiner phrase = determiner + noun phrase  
 b. noun phrase = noun + preposition phrase  
 c. preposition phrase = preposition + determiner phrase

For this reason, embedding is taken to signal recursion, underlying much of the debate between Everett (2005, 2009) and Nevins *et al.* (2009) on recursion as a

defining property of natural language. It will be recalled that Everett (2005) observed that the Amazonian language Pirahã lacks embedding, which was contested by Nevins *et al.* (2009) but upheld by Everett (2009). In the original 2005 article, the connection of embedding and recursion was not made (except in a comment by Michael Tomasello), but in his 2009 reply to Nevins *et al.*, the absence of recursion in the grammar of Pirahã was made the centerpiece of Everett's claim. At the background of this discussion is the hypothesis of Hauser *et al.* (2002) that recursion is the defining property of the faculty of language.

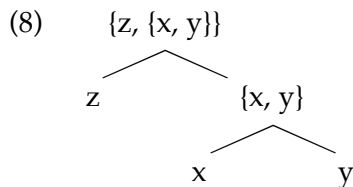
Meanwhile it should be clear that the model of grammar as understood in current minimalism no longer involves rewrite rules of the type in (4)/(5). Instead, structure is created by a single operation Merge, which Chomsky (1995) defines as taking two elements and combining them in a set:

- (6) *Merge*
  - i. select  $x$
  - ii. select  $y$
  - iii. create  $\{x, y\}$

Merge may be (re)written as a rewrite rule:

- (7) *Merge*  
 $x, y = \{x, y\}$

It is (mostly implicitly) understood that the next time Merge runs, it takes the output of the previous run as part of its input. Thus,  $x$  or  $y$  on the left-hand side in (7) may be  $\{x, y\}$ , yielding recursion. For example, if the next run combines  $\{x, y\}$  with  $z$ , the structure in (8) results.



(8) is a familiar binary branching phrase structure (where  $\{x, y\}$  may be written as the projection of  $x$ ,  $X'$ , and  $\{z, \{x, y\}\}$  as the maximal projection  $XP$ ). Such molecular structures are arguably the building blocks of all natural language phrases (including those in (3)), suggesting that every phrase consisting of more than two elements is the output of a recursive procedure (Nevins *et al.* 2009: 366).

However, just like Calvin's band of duplicates, the phrase structure in (8) may be generated by an iterative procedure as well. For instance, we may think of the structure-building process as a transfer procedure, moving elements from a resource (the lexicon, or a subset of the lexicon, or some other collection of elements called *Numeration* in Chomsky 1995) to a workspace (the structure under construction).

This procedure is spelled out in (9):

- (9) *Transfer*
- i. Move  $y$  from the Numeration to the workspace.
  - ii. Move  $x$  from the Numeration to the workspace.
  - iii. Move  $z$  from the Numeration to the workspace.

One could then define a constituent as the state of the workspace after each step.

An operation like *Transfer* is clearly iterative. It simply ticks off members of the resource set (the Numeration), and yet it generates the same structures as the recursive procedure that yields (8), essentially because the workspace is defined as the list of elements ticked off, and constituents are defined derivationally as subsequent stages of that list. The example illustrates that we cannot take for granted that binary branching structure involves a recursive procedure.

The point of the example is not to argue that *Merge* does not exist or that *Transfer* is superior to *Merge*. The point is that *Merge* and *Transfer* are equivalent in important respects, and though *Merge* may be a useful shorthand for *Transfer*, recursion should be identified in terms of the process, not its notation.

More generally, we cannot decide that a language (or natural language) ‘is recursive’ by simply looking at its structures. We have to know about the procedure by which these structures are derived (see also Everett 2009: 438).

## 2. Layered Derivations

There is ample reason to believe that the procedure generating clauses must be simple. Constituency tests generally yield the result that structures are binary branching (Kayne 1994), and structure-to-order conversion is not random but more or less automatic (Zwart 2011). This suggests a simple, stepwise procedure involving no more than two elements in each step, creating structure incrementally in a way that can easily be tracked by interpretive components of the mind/brain.

I have argued elsewhere that *transfer* (or a similar procedure) is in fact simpler than *Merge*, in that it identifies in each step a single element to be manipulated (transferred), while keeping the destination of transfer constant throughout the derivation (i.e. there is a single workspace). This yields binary branching structures with asymmetric sister pairs, essentially a nest of ordered pairs which can be linearized straightforwardly (along the lines of Fortuny 2008; see Zwart 2009, 2011).

One can think of *transfer* as an iterative procedure selecting (or identifying) one element at a time from the Numeration, creating a record of elements identified. As argued in Zwart (2009), this procedure can be viewed as working from the top down, splitting the Numeration in an element identified and a residue set. For example, (8) can be derived by starting from the Numeration in (10), via the steps in (11).



(10) Numeration = {*x*, *y*, *z*}

(11) *Transfer*

- i. split off *z* yielding ⟨*z*, {*x*, *y*}⟩
- ii. split off *x* yielding ⟨*z*, ⟨*x*, {*y*}⟩⟩
- iii. split off *y* yielding ⟨*z*, ⟨*x*, ⟨*y*, {∅}⟩⟩⟩

It is easy to see that transfer can yield embedding structures without recursion (see also Christiansen 1994, referred to in Parker 2006: 184). For example, (3a) can be derived from the Numeration in (12) by iterative splitting:

(12) Numeration = {John, thinks, that, Mary, said, that, it, was, raining}

Embedding, then, does not betray recursion if the structure building procedure is as simple as suggested here.

However, it is clear that not all sequences can be derived by iteratively splitting off the elements in the sequence from a Numeration containing them. Essentially, every complex non-complement (e.g., a subject) must be included in the Numeration as an atom. To see this, consider the derivation of (13).

(13) The man kicked the ball.

Constituency tests show that *the man* and *kicked the ball* in (13) are constituents (Chomsky 1964: 983). For example, *kick the ball* can be isolated in VP-fronting constructions like (14).

(14) The man said he would kick the ball, and *kick the ball* he did.

Likewise, *the man* can be isolated in the backgrounding construction in (15).

(15) He kicked the ball, *the man*.

But the sequence *man kick(ed) the ball* cannot be identified as a constituent by any known test.

The procedure generating (13), then, must yield a structure in which *the* and *man* are joined together before *the man* is joined with the rest of the clause. Iterative transfer starting from the Numeration in (16) could not yield this result, as can be seen after the first application of split in (17a).

(16) Numeration = {the, man, kicked, the ball}

(17) *Transfer*

- i. split off *the* yielding ⟨the, {man, kicked, the ball}⟩

It must be, then, that *the man* is included as a single item in the Numeration under-

lying (13):

(18) Numeration = {[the man], kicked, the ball}

Transfer then yields the correct result in terms of constituent structure:

(19) *Transfer*

- i. split off *the man* yielding ⟨[the man], {kicked, the, ball}⟩

It follows that *the man* must have been created in a separate derivation (or derivational layer), with the Numeration in (20) and the iterative transfer procedure in (21).

(20) Numeration = {the, man}

(21) *Transfer*

- i. split off *the* yielding ⟨the, {man}⟩
- ii. split off *man* yielding ⟨the, ⟨man, {∅}⟩⟩

The output of this derivation is included in the Numeration for the next derivation (i.e. (18)).

It should be clear now that the derivation as a whole, including the layers (20)/(21) and (18)/(19), is recursive: The output of the procedure (20)/(21) is part of the input (the Numeration) of the same procedure in (18)/(19). If this is the correct approach, recursion is not evidenced (necessarily) by embedding, but it is evidenced (necessarily) by left branch embedding (sometimes called ‘left-tail recursion’). (An alternative derivation starting from the Numeration in (16) merges *the* and *man* and *kicked, the*, and *ball* in two parallel derivations with outputs of the parallel derivations to be joined afterwards. In this alternative, a Numeration is associated with multiple workspaces, creating added complexity.)

### 3. Center-Embedding

Recursion in language is also typically illustrated by examples like (22), involving center-embedding.

(22) The dog the cat bit barked.

Here, *the cat bit* is a restrictive relative clause modifying *the dog*. The constituency of (22) is as indicated in (23):

(23) [ [ the dog [ the cat bit ] ] barked ]

In (22)/(23), we again see a complex subject (*the dog the cat bit*) which must have been derived in a separate derivational layer. The structure of (22)/(23), then, is recursive already for the same reason that the structure of (13) is. Center-embedding reduces to left branch embedding (and ‘nested recursion’ is in a natural class with ‘left-tail recursion’, if we are right).

It has been observed that center-embedding cannot be performed indefinitely, unlike right-branch embedding (Yngve 1961). Consider the triple embedding in (24), contrasting markedly with the triple embedding in (3a).

(24) [ [ the dog [ the cat [ the man kicked ] bit ] ] barked ]

It seems, then, that recursion (as understood here) comes with a cost (perhaps in memory load, as Yngve suggested). Much better again are (25) and (26):

(25) The dog that the cat bit that the man kicked barked.

(26) The dog that bit the man that kicked the cat barked.

The difference may be accounted for by the fact that (25) and (26) contain fewer derivational layers (due to the right-branch embedding) and hence less recursion.

#### 4. Interface Effects between Derivational Layers

If recursion in language is correctly defined as the interaction between derivational layers, it is easy to identify other phenomena that signal recursion in this sense. Assuming the model of grammar entertained in minimalism, each derivation (sequence of operations Merge) feeds the interface components dealing with sound and meaning, and hence we may expect the output of a derivational layer to show idiosyncratic sound/meaning properties. Moreover, we expect the output of a separate derivational layer to behave as an atom in the context of the next derivational layer: Constituents of the output of a separate derivation are not themselves in the numeration for the next derivation, and therefore cannot be merged individually in the context of that next derivation. This, I believe, derives a wide range of opacity effects, as first observed in Toyoshima (1997).

Among the interface effects affecting derivational layer outputs are (cf. Zwart 2009):

- (27) a. morphological effects (incorporation, conflation, fusion)
- b. atomization (opacity)
- c. idiosyncratic meaning (idioms)
- d. linearization effects (idiosyncratic order, template effects)
- e. categorization effects (reanalysis)

The effects in (27) all show a mix of syntactic and lexical properties. For instance, complex verbs (e.g., involving a transitivizing morpheme, possibly covert, like the ‘little *v*’ of Chomsky 1995) are clearly structured, but also clearly lexical (in obeying lexical integrity, for instance). The mixed properties are accounted for if such elements are created in a separate derivational layer, and pass through the interface components before being enlisted in the Numeration for the next derivational layer. If this is the correct approach, then every structured lexical item that is used in a larger syntactic context betrays recursion in the sense understood here.

In view of this, I would like to define ‘lexical’ in relative terms, i.e. in the context of a derivational layer, as in (28). (Note that under this definition, complex subjects are lexical items.)

(28) *Lexical*

*x* is a lexical item for derivation D of numeration N if *x* is included in N as a single item.

## 5. Recursion in Pirahã

The question now arises to what extent a language like Pirahã shows signs of recursion in the sense of derivational layering. I will try to answer the question, using data from Everett (1986) that are uncontested (and confirmed by Dan Everett, p.c.), by looking for (i) complex subjects and (ii) structured lexical items (in the sense of (28)) showing interface effects.

### 5.1. Complex Subjects

I believe it is uncontested that Pirahã does have phrases, and therefore, if Pirahã lacked recursion (in our sense), we would expect phrases to show up only in complement position, not in subject or adjunct position. Indeed, the large majority of the subjects in the examples of Everett (1986) appear to be single-word items (though they may be nominalizations or compounds, for which see below). But even so, several examples of complex subjects can be found in the examples (numbers refer to the example numbers in Everett 1986):

(29) *Complex subjects in Pirahã* (not including nominalizations/compounds)

a. **Xipoógi hoáoíi** hi xaagá. (22)  
*Xipoógi* *shotgun* *3* *be*  
 ‘That is Xipoógi’s shotgun.’

b. **Xoogiái hi xapisí** biga aí big-á. (43)  
*Xoogiái* *3* *arm* *thick* *be* *thick-EMPH*  
 ‘Xoogiái’s arm is thick (i.e. strong), very strong.’

- c. **giopaí gáihí** kapióxió xigiábií. (85)  
*dog that other like*  
 ‘That dog looks like another (dog).’
- d. **ti xahaigí** gáihí. (196a)  
*1 brother that*  
 ‘That (one) is my brother.’
- e. **kaolí xahaigí** gáihí. (196b)  
*who brother that*  
 ‘Whose brother is that?’
- f. **baaí xaíbaí** pii ap-ái-p-i pii bo-ó gai kob-á.  
*wild pig many water enter-ATELIC-IMPF-PROX water up-LOC DEM see-REM*  
 ‘A herd of pigs is entering the water upriver, look!’ (277)
- g. **Xoogíái hi go-ó hoasígikoí bíib-i híx** hoasígikoí  
*Xoogíái 3 WH-OBL lead shot send-PROX C lead shot*  
 koab-áo-b-í-i. (282)  
*run.out-TELIC-PERF-PROX-EVID*  
 ‘The lead shot which Xoogíái sent ran out.’

Of these, (29a,d,e) are copular constructions, which might allow another analysis in which the boldface noun phrase is a predicate. I have no idea about the plausibility of either analysis. Everett (2009: 419) comments on the copula-less type in (29d) that it is not a clause but just a string of nouns. But this does not affect the argument, unless the string looks like (30a) rather than (30b).

- (30) a. Me, brother, that.  
 b. [My brother], that.

If it is (30b), then, assuming iterative Merge and layered derivations, *my brother* must still be the output of a separate derivation.

Example (29g) involves a relative clause where it looks like the boldface material is not actually the subject, but a pre-posed topic, perhaps juxtaposed to the main clause. In that case, the boldface material would constitute a complex adjunct, which would have to be the output of a separate derivation under the assumptions entertained here.

## 5.2. Complex Lexical Items

Pirahã verbal morphology is fairly complex, and described in Everett (1986: 288–289) as templatic, featuring 18 slots following the verb root. None of these slots are occupied by inflectional morphemes. The categories represented include aspect, negation,

interrogativity, deixis, mood, but not tense. The complete template as proposed by Everett (1986) is in (31).

(31) *Pirahã verbal template* (all following the root)

- |                  |                  |                                |
|------------------|------------------|--------------------------------|
| 1. incorporation | 7. continuative  | 13. frustrative                |
| 2. duration      | 8. interrogative | 14. intensive                  |
| 3. telicity      | 9. ingressive    | 15. emphatic                   |
| 4. perfectivity  | 10. deictic      | 16. complementizer/nominalizer |
| 5. desiderative  | 11. iterative    | 17. evidential                 |
| 6. negation      | 12. certainty    | 18. result                     |

The first slot behind the root is reserved for incorporation (mainly of verb roots), “an extremely productive method of forming new verbs” (Everett 1986: 300–301).

Examples of complex verbs derived via incorporation are:

(32) *Pirahã incorporation*

- |    |                     |        |
|----|---------------------|--------|
| a. | xab op              | (388a) |
|    | <i>turn go</i>      |        |
|    | ‘return, arrive’    |        |
| b. | xiga hoag           | (388b) |
|    | <i>take come</i>    |        |
|    | ‘bring’             |        |
| c. | xig ab op           | (388c) |
|    | <i>take turn go</i> |        |
|    | ‘bring back’        |        |

As Everett (1986: 301) notes, the complex verb is treated as a single unit: Neither root can take any affixes, and “suffixes are added to the entire stem as one element”. I take this to entail that the incorporated verbs in Pirahã are prototypical structured lexical items as discussed above, i.e. outputs of separate derivational layers.

The etymology of the affixes is not generally elucidated in Everett (1986), but in a few cases he notes that the verbal suffixes are grammaticalized lexical items. For example, the evidential suffix *-xáagahá* ‘OBSERVATION’ is analyzed as involving *xaagá* ‘be’ and *-há* ‘COMPLETE CERTAINTY’ (Everett 1986: 298).

(33) *-xáagahá* < *xaagá* + *-há*.  
 OBSERVATION be COMPLETE CERTAINTY

This suggests that the verbal complex is the result of conflation as discussed in Hale & Keyser (2002), a typical syntactic process creating lexical items.

The complementizer/nominalizer slot 16 in (31) is occupied by *-sai*, a very productive morpheme for nominalizing verb phrases (Everett 1986: 277f.) (there is

also another nominalizer, *si*). One of its uses is quotative, affixed to a verb of saying, suggesting that it may be an embedding complementizer (cf. Nevins *et al.* 2009: 382).

- (34) *ti gái-sai kó'oi hi kaháp-íí.* (Everett 2005: 629)  
 1 say-NOM Kó'oi 3 leave-INTENTION  
 'I said that Kó'oi intends to leave.'

But Everett (2009: 418) takes *-sai* to be a marker of old information, allowing a juxtaposition rather than subordination analysis of the type in (34), where the first clause represents information that is familiar from the discourse setting (presumably akin to backgrounded quotatives in English).

Far more important to the discussion at hand, it seems to me, is the observation that *sai* turns verb phrases into nouns, the kind of process that betrays derivational layering (27e). Examples abound:

- (35) *Pirahã nominalization*
- a. *kahái kai-sai.* (80b)  
*arrow make-NOM*  
 'arrow making, arrow maker'
  - b. ***xíí kai-sai*** *hiaba.* (200)  
*thing make-NOM NEG*  
 'This is not a factory.'
  - c. *agaoakait-i-sai* (243)  
*canoe bore-LINK-NOM*  
 'canoe-boring-thing'
  - d. ***tiobáhai hóoi ai-sai*** *xabahíoxoi.* (262)  
*child bow make-NOM incorrect*  
 'Children's bow making is incorrect.'
  - e. ***ko kab-i-si*** *baósaápisi bag-áo-b-á-há.* (280)  
*eye NEG-LINK-NOM hammock sell-TELIC-PERF-DIST-EVID*  
 'The man without eyes (blind one) sold the hammock.'
  - f. *gahió pi-ó xabaip-i-sai* (288)  
*airplane water-LOC sit-LINK-NOM*  
 'Hydroplane'

In a layered derivation approach, nominalizations are perfectly regular syntactic constructions, merged with a nominalizing morpheme, and then turned over to the interfaces, acquiring idiosyncratic sound/meaning properties. At the interface, the output receives a new categorial feature (N), not transparently derived from any of

its constituent parts. The resulting unit can be included as a single item in the Numeration for a next derivational layer.

In this context, it is important to note that the nominalizing process is instrumental in creating Pirahã names:

All names for people are derived from verbal constructions, animal names, nominal phrases, etc. In about 90% of these cases, *-si* occurs optionally in morpheme final position, as though marking a change in the basic reference or function. (Everett 1986: 279–280)

A final category that suggests derivational layering in the grammar of Pirahã involves compounds. Everett (1986) lists numerous compounds, but Everett (2009: 423–424) appears to argue that these are not really compounds, admitting that:

If there were compounding in Pirahã, this would be clear evidence for recursion. (Everett 2009: 423)

Let us first look at the examples:

(36) *Compounds in Pirahã*

- a. **xagí gahióo** xogíí ái-xi-xi      pii xigiábií.      (86)  
*path airplane big be-EMPH-EMPH water like*  
 ‘The **airstrip** is big, like a river.’
- b. xogaogíí < xogaí + ogíí      (389)  
*big field field big*
- c. xabagisoixaoxisai < xabagi + soixaoxisai      (477)  
*saw toucan beak*
- d. xapaítoii < xapaí + toii      (478)  
*ladder foot handle*
- e. pigáíá < pi + gáíá      (481)  
*scissors thorn crooked*
- f. kaogíái < kao + ogíái      (482)  
*[kind of bass] mouth big*

These all seem clear cases of compounds, in most cases acquiring the idiosyncratic (non-compositional or metaphoric) meaning suggestive of derivational layering.

Everett (1986: 322) describes his grounds for classifying formations like those in (36) as compounds as semantic: The non-compositional meaning suggests they are lexical items rather than phrases. Later, Everett (2009: 423–424) withdraws the semantic argument and details the prosodic properties of the formations in (36), suggesting to him that they are not compounds. With the remarks of Everett (1986: 322) in mind (see below), we may conclude that he considers (36) not to involve compounds but (syntactic) phrases.



The criterion to classify the examples to follow as compound words rather than merely phrasal constructions is semantic. (Everett 1986: 322)

In the layered derivations approach, there is no fundamental distinction between ‘compound words’ and ‘phrasal constructions’. What is relevant is the complexity of the string (suggesting derivation output status) and its behavior as a single item in the context of a (subsequent) derivation. The idiosyncratic meaning merely provides an additional argument for these elements’ derivational history, regardless their status as words or phrases.

In this context we may also point to complex locatives in Pirahã, suggestive of the kind of fusion that we expect to occur at the interface between derivational layers (cf. (27a)).

(37) *Complex locatives in Pirahã*

- |    |          |                   |       |
|----|----------|-------------------|-------|
| a. | xoí      | ‘jungle’          | (326) |
| b. | xo-ó     | ‘in the jungle’   |       |
| c. | xo-ó-xio | ‘into the jungle’ |       |

## 6. Conclusion

In conclusion, if recursion is identified in terms of derivational layering, as proposed here, then it seems clear that the grammar of Pirahã is recursive. Both complex subjects and complex lexical items are attested, some in average quantity, some in abundance.

This raises the further question what the observations in Everett (2005), substantiated in (2009), about the absence of embedding imply. Clearly, the wide-ranging implications having to do with the nature of the faculty of language vanish, but the original conclusions of Everett (2005) were considerably less bold and may still be valuable. One suggestion to make, within the model of grammar considered here, is that while interaction among derivations is unaffected by whatever cultural constraints are at play, the size of the Numeration is. Recall from section 2 that embedding structures like (3) can be derived by iterated Merge, starting from a large enough Numeration. Perhaps the ‘immediacy of experience principle’ that Everett (2005) suggested to capture the cultural constraints on Pirahã grammar (and cognition) constrains the Numeration in ways that virtually preclude ordinary right-branch embedding.

## References

Chomsky, Noam. 1964. The logical basis of linguistic theory. In Horace G. Lunt (ed.),

- Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass., August 27–31, 1962*, 914–1008. The Hague: Mouton.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Christiansen, Morten H. 1994. Infinite languages, finite minds: Connectionism, learning and linguistic structure. Edinburgh: University of Edinburgh dissertation.
- Everett, Daniel. 1986. Pirahã. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 1, 200–325. Berlin: Mouton de Gruyter.
- Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46, 620–646.
- Everett, Daniel. 2009. Pirahã culture and grammar: A response to some criticisms. *Language* 85, 405–442.
- Fortuny, Jordi. 2008. *The Emergence of Structure in Syntax*. Amsterdam: John Benjamins.
- Hale, Ken & Samuel J. Keyser. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- 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* 198, 1569–1579.
- Kayne, Richard S. 1994. *Connectedness and Binary Branching*. Dordrecht: Foris.
- Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85, 255–404.
- Parker, Anna R. 2006. Evolution as a constraint on theories of syntax: The case against minimalism. Edinburgh: University of Edinburgh dissertation.
- Toyoshima, Takashi. 1997. Derivational CED. *WCCFL* 15, 505–519.
- Yngve, Victor. 1961. The depth hypothesis. *Proceedings of Symposia in Applied Mathematics* 12, 130–138.
- Zwart, Jan-Wouter. 2009. Prospects for top-down derivation. *Catalan Journal of Linguistics* 8, 161–187.
- Zwart, Jan-Wouter. 2011. Structure and order: Asymmetric merge. In Cedric Boeckx (ed.), *Oxford Handbook of Linguistic Minimalism*, 96–118. Oxford: Oxford University Press.

Jan-Wouter Zwart  
 Rijksuniversiteit Groningen  
 Center for Language and Cognition Groningen  
 P.O. Box 716  
 9700 AS Groningen  
 The Netherlands  
[c.j.w.zwart@rug.nl](mailto:c.j.w.zwart@rug.nl)

# The Acquisition of Recursion: How Formalism Articulates the Child's Path

Tom Roeper

We distinguish three kinds of recursion: Direct Recursion (which delivers a 'conjunction' reading), Indirect Recursion, and Generalized Transformations. The essential argument is that Direct Recursion captures the first stage of each recursive structure. Acquisition evidence will then be provided from both naturalistic data and experimentation that adjectives, possessives, verbal compounds, and sentence complements all point to conjunction as the first stage. Then it will be argued that Indirect Recursion captures the Strong Minimalist Thesis, which allows periodic Transfer and interpretation. Why is recursion delayed and not immediate? It is argued that an interpretation of Generalized Transformations in the spirit of Tree Adjoining Grammar offers a route to explanation. A labeling algorithm combines with Generalized Transformations to provide different labels for recursive structures projection. Recursion is then achieved by substitution of a recursive node for a simple node. One simple case is to substitute a Maximal Projection for a simple non-branching lexical node. A more complex case — essential to acquisition — is to substitute a category for a lexical string. Consequently, a computational 'psychological reality' can be attributed to explain why recursion requires an extra step for the addition of each recursive construction on the acquisition path.

*Keywords:* direct recursion; generalized transformation; indirect recursion; interfaces; phases; strong minimalist thesis

## 1. Introduction

Why would a child who can say *wagon-puller* not be able to understand *wagon-*

---

Thanks to many partners in this research: William Snyder, Bart Hollebrandse, Jill de Villiers, Chloe Gu, Maxi Limbach, Kate Hobbs, Misato Hiraga, and Ed Matthei have explored five major constructions, via the study of naturalistic data and a variety of different experiments. Bart Hollebrandse has explored both the theoretical side and been central in experimental work. William Snyder was the original catalyst for much of this research. Conversations with Noam Chomsky, Tecumseh Fitch, Uli Sauerland, Rajesh Bhatt, and Robert Frank were also helpful, and especially written comments from Robert Berwick. Audiences at DGfS, CHLASC, Rio de Janeiro, Tromsø, Trondheim, Tokyo, Harvard, Essex, have been just the recent ones exposed to this evolving research — and I thank them all for helpful commentary. This work was supported by NSF grant to Margaret Speas, Thomas Roeper, Jill deVilliers, and Jay Garfield.



*puller-maker*? Why is language-specific recursion not immediate? Our goal is to articulate the acquisition challenge, review the relevant evidence, and imagine why there is an acquisition path for recursion. The evidence leads to a rather tight grammatical edifice, which, however, is full of theoretical and empirical weak points that deserve further research. Why such a strategy? It is really the strategy of linguistic theory in general. Weak points — like question-mark data — can be strengthened by other branches in a logical and empirical, hence persuasive, edifice once general properties are identified. If we follow the acquisition path of several kinds of recursion, their faint light can become a strong beam when seen together.

In a similar vein, a number of formal alternatives<sup>1</sup> become much sharper when we attach them to empirical phenomena, even if the data might seem open to many interpretations.<sup>2</sup> The major task at hand, ultimately, is to build transparent interfaces between structure and interpretation and, as well, an interface between a theoretical account and the actual time course of acquisition. Yet, like the evolving notation of theoretical linguistics, these are proposals about how to build a notation that responds to both the facts of recursion and the acquisition path, neither of which is fully evident. Therefore we include pilot data and naturalistic data which might seem insufficient for traditional psychological experimentation, but which, in light of powerful theoretical proposals, become legitimate reference points in the interaction between theory and empirical data. This first ‘fieldwork’ stage of acquisition needs a recognized legitimacy as an important background to detailed work, much as rough awareness of language variation in unusual languages tempers broad claims about UG.

First, let us distinguish between a completely universal form of recursion, namely Merge, and language-specific forms. Merge is a binary recursive operation that is invoked as soon as more than two words are combined.<sup>3</sup> Therefore all languages with 3 word combinations are examples of recursion over

---

<sup>1</sup> This is written by someone who is by no means well-trained in mathematical or formal notation. See Tomalin (2007) and Lobina (2010) for formal discussions that articulate some of the distinctions between the formal and the empirical approaches to notation. Nonetheless, the argument does build upon, in a broad way, the implicit biolinguistic philosophy that language formalism should be built straight from empirical data — much like the double-helix model in biology — rather than adhering to theorems that logic or mathematics have derived within their own systems. Thus Chomsky (2010) argues, as I understand it, that linguistic formalism should use concepts from set theory without being built from its theorems.

<sup>2</sup> In particular, one might seek to reduce the arguments to processing claims as proposed by Berwick & Weinberg (1986) and to various proposals that claim that sentences are first parsed as a series of conjunctions, then interpreted in a second pass by imposing dominance relations (see Stabler, to appear). Whatever the merit of these claims about hearer parsing, if the phenomena can be resolved into time-free structural grammatical representation, then we take this to be a superior account. Some of this structure may be a reflection of ‘externalization’ demands that are tighter than grammatical demands, but a unified account would definitely help computational efficiency at any level (see Berwick & Chomsky 2011).

<sup>3</sup> It is exactly a binary, not a ternary operation, following Chomsky (2010) who argues that theories of sideward movement and multi-dominance constitute ternary merge, a deviation from this essentially biological claim about how core grammar works, but perhaps another reflection of Direct Recursion.

two binary acts of Merge. It is possible to imagine a three-term concatenation without a binary substructure, but empirical arguments exist to demonstrate that this is not the case for humans in structure-building beyond conjunctive relations, which we will elaborate in what follows.<sup>4</sup> The presence of recursive Merge means that all languages must be recursive in a fundamental sense, just as Hauser *et al.* (2002) have claimed, which constitutes a strong biological claim.

Nevertheless, not only the empirical question, but also the formal question remains alive: Recursion may not be captured by a single formalism or be represented as a single object in the brain.<sup>5</sup> In the *Prism of Grammar* (Roeper 2007), I argued that the principles of grammar are a model for how we envision other mental operations. If true, then other analogies should be available as well. Stereoscopy, the integration of information from two sources, is one concept in science, but it applies to both eyes and ears. Nevertheless, the purposes and neurology of eyes and ears are quite different, and therefore it is obvious that they must be separately represented in the neurology of the brain and its consequent informational representations. There is no single stereoscopy center in the brain. It is possible that we need to look at recursion in the same manner. In other words, our ultimate understanding will involve coordinated representations in both grammatical and biolinguistic terms, which may be conceptually and biologically distinct from apparent forms of ‘visual’ recursion (see Jackendoff 2010, Berwick & Chomsky 2011).<sup>6</sup>

We will focus upon three representations for building recursive structure (defined below): Direct Merge, Indirect Merge, and Generalized Transformations (GTs) as realized in an adaptation of Tree-Adjoining Grammar (TAG). Direct Merge allows a category to generate itself, while indirect Merge introduces an identical category only through another non-identical one. GTs combine pre-existing structures. Related, non-recursive forms include iterativity (as in ‘*very, very big*’) and Concord (‘I don’t want *any* food at *any* time for *any* reason’). It is an interesting question, particularly from a biological perspective, whether there are deep connections among recursion, iterativity, and concord — but they lie beyond what we can approach here.

A critical ingredient in our account is the interface with interpretation,<sup>7</sup>

---

<sup>4</sup> One might, in fact, argue that Direct Recursion and conjunction constitute a non-grammatical interface with phase-based grammar. This would explain why conjunctions create islands and have links to very challenging forms of across-the-board movement and gapping. In a sense, then, language acquisition begins when children re-analyze conjoined representations. Then conjunction would belong to Primary Linguistic Data representations (Chomsky 1965) whose representational characteristics deviate from what we find in Final Grammars, except in marginal constructions where they reappear. Bob Berwick (p.c.) points out that Miller & Chomsky’s (1963) explanation of center-embedding essentially claims that conjunction is the default to which the grammar returns. It is significant, then, if acquisition reveals the same default in a host of different constructions.

<sup>5</sup> Bob Berwick (p.c.) observes that this is essentially the position of David Marr on the status of representations for vision in the brain.

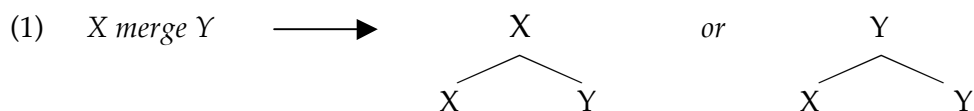
<sup>6</sup> We have no view of this matter, but Noam Chomsky (p.c.) and Bob Berwick (p.c.) argue that visual and mathematical abilities are parasitic on grammatical recursion.

<sup>7</sup> See Roeper (2009) for discussion of the distinction between an interaction and an interface. Mechanical interfaces are like cogs in a wheel, while interactions may not always have a principled basis.

which, we argue, is linked to Indirect Merge in an important, putatively innate, way. Ultimately the ideal notation for an interface should be transparent for both syntax and semantics. The critical biological claim is that there is a strict interface between points of recursion and points of interpretation. One could imagine that an organism could have both capacities, but lack the interface.<sup>8</sup>

## 2. Merge and Labeling Algorithms

Merge is the putative universal form of an operation that underlies any form of syntactic hierarchical structure, as in (1).<sup>9</sup> Although a set may be defined without a label or ordering, a signal feature of Merge lies in the fact that human languages always assign a Label to every Merge.<sup>10</sup>



A label must be chosen reflecting the dominance of either the right or the left branch — or possibly a more complex choice; see Chomsky’s (1995, 2008) discussion of labeling algorithms.<sup>11</sup> Hornstein (2009) has suggested that it is the combination of Merge and labeling which may define human grammar as distinct from animal constructs. We take the argument one step further in arguing that the connection between recursion and the Strong Minimalist Thesis (SMT, see below), which argues that certain nodes represent ‘phases’ which carry an interface with interpretation.

### 2.1. Direct and Indirect Recursion

An initial distinction between direct and indirect recursion can be made in terms

<sup>8</sup> Roeper (1978) argued that both hierarchy and node labels could have other origins but the innate property of language is to link them in a fixed way. Hauser (2008) argued that animals might have all abilities, but lack only the interfaces. In ongoing work, Ray Jackendoff argues that it is the connection to words that is unique to grammar. This view will reduce to the categorical feature that words carry and thus relate to the labeling claim.

Left unarticulated is how to interpret the common idea that “recursion is a general cognitive capacity” (Ray Jackendoff’s ongoing work but also Everett, to appear). The term *general* is where we must be careful. We might say that ‘motion’ is a general capacity of all muscular organisms, but the claim would have little biological force since it is obvious that the mechanisms for motion of eyelids and legs are so differently represented in different organisms, or different parts of one organism. This is analogous to our argument that there is no Stereoscopy Center in the brain, but rather the principle is independently represented in eyes and ears.

<sup>9</sup> See Roeper (2003) for discussion of the interaction between successive forms of Merge and compositionality for DP.

<sup>10</sup> A question, of course, arises about how linearization occurs and whether it belongs to a process of externalization, as Chomsky (2010) has suggested. If such a reframing occurs, the role of order and labeling in the definition of the externalization interface would leave these claims unchanged, I believe.

<sup>11</sup> See Roeper & Pérez-Leroux (2010) for discussion.

of phrase-structure rules (Snyder & Roeper 2004). Direct recursion is where a category reproduces itself and characteristically produces a conjunctive reading:

- (2) Direct Recursion:  $X \rightarrow Y(X)$   
 $NP \rightarrow NP ((\text{and}) NP)$

This will produce potentially infinite sentences like:

- (3) John, Bill, Fred, and Susan arrived.

It has a critical feature: There is no significant semantic ordering among the elements. They are parallel and interchangeable:

- (4) Bill, Susan, John, and Fred arrived.

It is applicable to any category, even below the lexical level:

- (5) a. in and around and over and under the structure  
 b. pre- and post-operative care

but does not participate in other aspects of grammar; for instance, there are no movement rules that allow extraction from conjunction (see Ross 1967):

- (6) \* how did he go in and \_\_ the structure  $\rightarrow$  'around'

It is, in a sense, at the margins of grammar, but it is also a mental ability that characterizes the first stage — and the default grammar — of children with respect to every category in the grammar, as we shall illustrate.

Bob Berwick (p.c.) makes the interesting suggestion that conjunction means one simply relaxes the 'completeness' condition on dominance and precedence such that not every two phrase markers must be in a dominance (or precedence) relation. In terms of *John, Bill, Fred, and Susan arrived*, this model says that the NPs bear no syntactic relation to one another.<sup>12</sup> In sum, conjunction carries no dominance relations, and therefore a basic tenet of grammar is not honored.

Such an account matches ours, but it is worth articulating that, if the child assumes no grammar here, then inference must supply much of the meaning, including implications. Sentences like *John got drunk and Bill got angry* allow a causative inference for adults, and perhaps for children too. It can be seen as part of the acquisition process that a child exchanges broad and unreliable inferences for syntactically guided compositional meanings.

By contrast, indirect recursion may (or may not) involve an interpretive step that changes meaning, as in the way that possessives are stacked:

<sup>12</sup> Conjunctive *and* inside grammar has other subtle consequences. The default notion of Conjunction should not be seen as co-extensive with the grammar of *and*. See Munn (1999) for arguments on the special status of first conjuncts. *and*-conjunction also interacts, for instance, with binding theory. Nor should default conjunction be identified with its use in logic. It is essentially a minimal, non-syntactic association that invites all kinds of inferences, much like root compounds are open to inference: *Elephant icebox* could be an icebox for elephants, or one that looks like an elephant or has an elephant picture on it.

(7) John's friend's father's student's essay

is quite different from:

(8) John's student's father's friend's essay

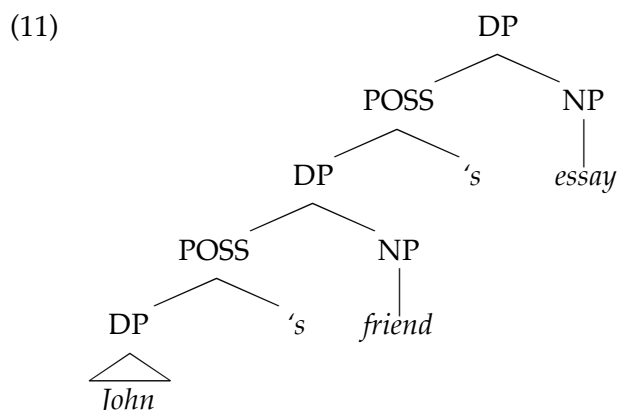
We can capture the difference by introducing the SMT (see Chomsky 2008, 2010):

(9) *Strong Minimalist Thesis (SMT)*  
Interpretation proceeds phase by phase.

Although it remains an open question just where phases occur, good arguments for CP, *v*P, PP, and DP as phases have been made (see van Hout *et al.* 2011). The recursion is indirect because another category is present:

(10) Indirect Recursion:    DP → (Determiner) NP  
                                  Determiner → {ARTicle POSSessive}  
                                  POSS → DP 's

The DP is repeated inside the possessive phrase, and therefore can generate another 's for *John's friend's essay*:



If interpretation occurs at each phase, the phase-assumption is critical,

(12) A DP is a phase.

which is a designated interpretive domain, as are CP, *v*Ps, and PPs.<sup>13</sup> In Chomsky's phase-theoretic formulation, Transfer sends a syntactic object SO to the semantic component, which maps it to the C-I interface; this SO is identified

<sup>13</sup> If indirect recursion occurs through an intermediate phase of a very different type, like a PP, it does not impose the same recursive interpretive demand:

(i) [<sub>DP</sub> *the* box [<sub>PP</sub> in [<sub>DP</sub> *the* corner]]]

A PP intervenes between two DP's. Here we are basically unaware that one Determiner (*the*) is inside another. Thus the possessive interpretation inside a possessive interpretation is where recursion has the effect we are after.



as a *phase*. Thus the SMT entails that “computation of expressions must be restricted to a single cyclic/compositional process with phases”. Or in full:

As noted, iterated Merge incorporates the effects of three of the EST/Y-model compositional cycles, while eliminating d- and s-structure. Still unaccounted for are the cyclic/compositional mappings to the phonetic and semantic interfaces. These too are incorporated, and the final internal level LF is eliminated, if at various stages of computation there are Transfer operations: one hands the SO already constructed to the phonological component, which maps it to the SM [sensorimotor] interface (“Spell-Out”); the other hands SO to the semantic component, which maps it to the C-I [conceptual-intentional] interface. Call these SOs *phases*. Thus SMT entails that computation of expressions must be restricted to a single cyclic/compositional process with phases. In the best case, the phases will be the same for both Transfer operations. (Chomsky 2008: 142)

As a strong constraint, it guides and constrains acquisition as well.

## 2.2. *Alternating Phase Constraint*

Boeckx (2009) argues for what we can call the Phase Alternation Constraint (see also Richards 2011 for a somewhat different implementation):

(13) *Phase Alternation Constraint: Interpretation must occur in alternating sequence*

Transfer takes place every other time Merge applies and yields the pattern:  
phase — non-phase — phase — non-phase

In each of the constructions above we find that this sequence is followed; every other time Merge applies yields the following pattern:

(14) a. {Head4 *Transfer*2, {Head3, {Head2 *Transfer*1, {Head1}}}}  
b. = [C *phase* [T [*v phase* [V]]]]

This leads to the following kinds of familiar alternations:

(15) Sentence: John thinks that Bill thinks that Fred...  
                  VP      CP          VP      CP

PP:          John’s knowledge of Bill’s knowledge of...  
                  DP          PP          DP          PP

(16) Possessive: John ‘s friend ‘s father ‘s car  
                  NP Poss NP Poss NP Poss NP

In sum, it is indirect recursion linked to the interpretive requirement (SMT) on phases that carries the weight of recursion as a pivotal grammatical device. We will now show how languages differ in where they allow indirect recursion, and then reveal a two-step acquisition path for each form of language-specific recursion.

### 2.3. Grammar Variation

A broad overview of how grammars may vary in recursion will help see the scope of the acquisition challenge. German (and most Germanic languages) allows a single pronominal genitive, limited to proper nouns:

- (17) a. Marias Haus  
       *Maria's house*  
       b. \*Marias Nachbars Freundins Haus  
       *Maria's neighbor's friend's house*

Therefore, the child needs to identify where in his language recursion occurs. In German we argue that the POSS directly dominates the lexical item 's and therefore does not dominate DP producing recursion.<sup>14</sup> This is the child's first assumption.

Among the major known recursion contrasts, where a single element but no recursive elements occur, we find the following: (i) single possessives, as in German *Marias Haus* 'Maria's house', (ii) single double verbs, as in English *come help*, (iii) single prenominal adjectives, as in French *pauvre enfant* 'poor child', (iv) single compounds, as in French *homme-grenouille* 'man frog', (v) and single complements, as in Pirahã:

- (18) a. Compounds:           Germanic languages → recursion  
                                   Romance languages → no recursion  
       b. Possessives:       English → recursive possessives (Saxon  
                                   German → no recursive possessives  
       c. Adjectives:       English → recursive prenominal adjectives  
                                   no recursive post-nominal adjectives  
                                   French → no recursive prenominal adjectives  
                                   recursive post-nominal adjectives  
       d. Serial verbs:      Bantu → recursion  
                                   English → no recursion  
       e. PP recursion:     English → recursion  
       f. Clausal recursion: Germanic, Romance → recursion  
                                   Sign Language, Pirahã → (disputed)  
                                   Walpiri, Teiwa → no recursion

One important challenge is to uncover exactly where recursion occurs in less studied languages around the world. Each will provide an acquisition challenge.

<sup>14</sup> The contrast between lexical and phrasal possessives in English may mirror the English/German contrast. Lexical possessives (ii) cannot be phrasal unlike 's (i), as in this contrast:

- (i) the man next to me's hat  
 (ii) the man next to my hat

For instance, examples like the following Pirahã are disputed examples of complementation (from Sauerland 2010):

- g. Pirahã → hi ob-áaxái kahaí kai-sai.  
 see/know-INTNS arrow make-NOMLZR  
 'He really knows how to make arrows.'

Apparent subordination could be an effect of coordination, nominalization, and/or intonation which has led Everett (to appear) to claim it could be formed by parataxis, hence conjunction, although evidence seems to be mounting against this view.<sup>15</sup>

### 3. Data for Direct Recursion: The Appearance of *and*

Our fundamental claim about the first stage for every recursive structure is this:

(19) *Direct Recursion is the Acquisition Default*

A child first analyzes adjacent identical structures as Direct Recursion with a Conjunctive reading.

The claim is distinct from but compatible with parsing claims that conjunction is a preferred parsing strategy.<sup>16</sup>

The first evidence of a conjunctive interpretation arises in naturalistic data where *and* is frequent and arises where one senses that adults might normally put a different conjunction, although *and* is open to many inferences for adults too. There are three from Adam at age three and a half and others randomly selected from a CHILDES search (cf. MacWhinney 2000):

- (20) adam30.cha:\*CHI: when I lived in a bunkhouse # and I saw a snake coming out.  
 adam30.cha:\*CHI: and my teeth and I bite em.  
 adam29.cha:\*CHI: I goin(g) to put back # and I got something for his face.  
 57.cha:\*CHI: now they are a [/] awake and I open the door!  
 20a.cha:\*CHI: I'm gonna do it and I can turn the page.  
 16b.cha:\*CHI: I'm a bunny and I eat you.

<sup>15</sup> The claim that it is not a form of conjunction is difficult to establish. However, Sauerland (2010) provides intricate and interesting arguments that the linking morpheme *sai* may be a subordinator, particularly when intonation variation is present. Moreover, the presence of verbs like *know* suggests strongly the subordination relation. Nonetheless, it is possible that the clause could be subordinated without being recursive, as Perfors *et al.* (2011) suggest. In that case, the critical evidence would be a combination of double subordination and semantic opacity of the kind achieved in the English experiment reported here with embedded propositions whose truth the speaker does not assume. Uli Sauerland (p.c.) also has evidence of opacity in showing that Pirahã speakers can construe a subordinated clause as carrying a false belief. Thus the evidence looks strong that real subordination is present in complementation. The next step is to carry out the experiment reported here that shows recursive subordination with opacity which 6-year-olds in English can comprehend.

<sup>16</sup> Pietroski (2011) argues that a form of conjunction underlies all adult semantic interpretation as well, but the semantic notion in logic may not coincide with the argument here.



### 3.2. *Prepositional Phrases*

Naturalistic evidence from CHILDES analyzed by Chloe Gu shows that children will treat PP's conjunctively and resist recursion (see Gu 2008).

- (25) Father: Up in the shelf in the closet  
 Child: yeah  
 Father: can you say that  
 Child: up in the shelf in the closet  
 Father: very good, up in the shelf in the closet in the kitchen, can you say that  
 Child: yeah, up in the # up in the # what  
 Father: up in the shelf in the closet in the kitchen  
 Child: up in the shelf in the # what  
 Father: closet  
 Child: in the closet in the kitchen  
 Father: in the jar up in the shelf? can you say that?  
 Child: I can't  
 Father: you can  
 Child: in the jar # say in the jar  
 Child: up in the shelf in the jar in the closet in the kitchen

Note that the PPs are now conjoined (in the shelf and in the jar), rather than recursively embedded (the shelf is not in the jar). It would be good to gather experimental evidence on this point. The experiment is easy to see: Put a box on a shelf and one on the floor, and a book in each. Then ask: "Show me the book in the box on the shelf". If children treat the question as conjoined, they will point to both the book in the box on the shelf and the one on the floor. As we will see, this response is found with possessives.

### 3.3. *Recursive Possessives*

Naturalistic acquisition data on recursive possessives suggests that they are difficult (see Roeper 2007 for more examples):

- (26) MOTHER: What's Daddy's Daddy's name?  
 SARAH: uh.  
 MOTHER: What's Daddy's Daddy's name?  
 SARAH: uh.  
 MOTHER: What is it?  
 What'd I tell you?  
 Arthur!  
 SARAH: Arthur! Dat my cousin.  
 MOTHER: Oh no, not your cousin Arthur.  
 Grampy's name is Arthur.  
 Daddy's Daddy's name is Arthur.  
 SARAH: (very deliberately) No, dat my cousin.

MOTHER: oh.  
 What's your cousin's Mumma's name?  
 What's Arthur's Mumma's name?  
 MOTHER: What's Pebbles-' momma's name?  
 SARAH: Wilma.  
 MOTHER: Wilma... yeah.  
 And what's Bam+Bam's daddy's name?  
 SARAH: Uh, Bam+Bam!  
 MOTHER: No, what's Bam+Bam's daddy's name?  
 SARAH: Fred!  
 MOTHER: No, Barney.  
 SARAH: Barney.  
 MOTHER: What's his mumma's name?  
 SARAH: She's right here.

Sarah is resisting a recursive understanding although all the pragmatic support and world-knowledge she needs is close at hand.

A long dialogue where a father tries to get a child to simply repeat a recursive possessive shows that the child understands the meaning, but converts the possessive into a single possessive with a compound (see Roeper 2007):

(27) FATHER: How about the Dukes of Hazard's boy's car?  
 CHILD: Yeah.  
 FATHER: What is it called?  
 CHILD: The boy's Dukes of Hazard car.  
 FATHER: No, not the boy's Dukes of Hazard.  
 It's the Dukes of Hazard's boys.  
 Can you say that? Dukes of Hazard's boy's car?  
 CHILD: The boys Dukes of Hazard car. (repeated several more times)

A 6-year-old, though, produces one with ellipsis (marked by the transcriber as possessive and not plural based on context):

(28) where's Toto's girl's \_\_\_\_

The child initially finds any way possible to resist the interpretation that recursion demands. The favored move is to convert a recursive sentence into conjunctions as data below indicate.

### 3.3.1. Possessives Explored

In a series of explorations by various students and colleagues we began to pursue the question experimentally. The first step is to invent a context where several options are available and equally plausible. The first was invented by Sarah Gentile (2003), who gave a child three pictures based on familiar Sesame Street characters, but no story was presented (adults were tested in the next study).

- (29) A. Picture of Cookie Monster  
 B. Picture of *Cookie Monster and his sister*  
 C. Picture of his sister  
 “Can you show me *Cookie Monster’s sister’s picture?*”

The results showed that about one third of the 3- to 4-year-olds took the conjunctive reading (*Cookie Monster’s and sister’s picture*) and chose Picture B.

In the next experiment by Maxi Limbach, children and L2 German speakers whose L1 has possessives but no recursion, were given a series of stories, like this one, where both options are equally attractive:

- (30) Context story example for screen setting:

*Jane has a nice blue bike and Jane’s father Gordon has a racing bike. When they do a tour together, they have another bike which they can ride together. Sam has a red bike and his father Paul has a silver bike.*

After a presentation of all bikes and actors (Fig. 1), the bikes were then shown in separate pictures and participants chose which fit “Jane’s father’s bike”.



Bikes: Racing tandem blue  
 father’s both Jane’s

Figure 1: Recursive and conjunctive options for recursive possessives

Subjects who were either native (NS) or non-native speakers (NNS) were involved: 25 American English-speakers and 23 German university L2 students. 26 children were divided into three age groups — nine 3-year-olds (average age: 3;7), eight 4-year-olds (average age: 4;5), and nine 5-year-olds (average age: 5;7). NNS adults gave a conjoined reading or dropped one of the possessives (38%, compared with 37% for the 5-year-olds). It is noteworthy that the 5-year-olds gave 22% conjoined readings, while the NNS adults gave only 8%, preferring 30% of the time to drop the first or second possessive. Here are overall results (see Limbach & Adone 2010 for further analysis):

	All	Correct	Middle drop	First drop	Random (unrelated)	Conjunctive	Confusion
5-y.o.	32	19 (59%)	3 (9%)	2 (6%)	0	7 (22%)	1 (3%)
4-y.o.	23	16 (70%)	1 (4%)	1 (4%)	0	4 (17%)	1 (4%)
3-y.o.	32	18 (56%)	6 (19%)	2 (6%)	0	3 (9%)	3 (9%)
Adult (NS)	109	90 (83%)	2 (2%)	11 (10%)	1 (1%)	5 (4%)	41 missing
Adult (NNS)	102	63 (62%)	10 (10%)	12 (12%)	9 (8%)	8 (8%)	36 missing

Table 1: Results

Moreover, L2 speakers of English persistently claim that recursive possessives are difficult, and triple recursion virtually impossible.<sup>19</sup>

Why should it be hard to go beyond the single possessive? We note that possessive is a form of case assignment<sup>20</sup> in many grammars and it bears thematic roles in nominalizations that have nothing to do with possession (*the enemy's destruction of the city*). Other thematic roles cannot expand by recursion. For instance we cannot expand Agent or Theme in that manner: \**The wall was built by the company by the government* meaning via the agency of the government the agency of the company caused the wall to be built. Recursive agency cannot attach to possessive agents either: \**The government's company's building of the wall*. Such a sentence could mean the company owned by the government but not the company caused by the government to build. Once again, capturing the semantic side of the interface is a critical challenge to both the theory and the acquisition process.

### 3.3.2. Japanese

Now we look at a pilot experiment on recursive possessives in Japanese where, for the first time, four level recursion has been explored by Fujimuri (2010). In Japanese we have a structure similar to English but marked by *no*:

- (31) a. John's brother's car. *English*  
 b. John no otouto no kuruma. *Japanese*  
*John 's brother 's car*

A simple set-up was matched by a picture sequence that allowed the relations to be easily kept in mind.

- (32) The story (told in Japanese):

*This girl is Mika and this is her dog. This boy is Mika's friend and his name is Kenta. This is Kenta's dog. This is Mika's brother and his name is Sho. And this is his dog. This is Sho's friend, Yuki and this is her dog. And look, everyone is holding a ball.*

Alongside the story, the pictures of all actors were shown:

---

<sup>19</sup> It is interesting that even among professional linguists for whom English is not native, who have intellectual understanding (by their own testimony) does not make them able to produce them in conversation.

<sup>20</sup> Pointed out to me by Uli Sauerland (p.c.).





Figure 2: Pictures for multiple possessives in Japanese

(33) *single possessive questions:*

1. What color is Mika's ball? — Orange.
2. What color is Kenta's flower? — Yellow.
3. What color is Sho's shirt? — Red.

*double possessive questions:*

4. What color is Mika's dog's ball? — Black.
5. What color are Mika's brother's shoes? — Yellow.
6. What color is Sho's friend's ball? — White.

*triple possessive questions:*

7. What color is Mika's friend's dog's ball? — Purple.
8. What color is Mika's brother's friend's flower? — Red.
9. What color is Sho's friend's dog's tail? — Black.

*quadruple possessive question:*

10. What color is Mika's brother's friend's dog's ball? — Yellow.

Table 2 summarizes the responses of the seven children for the 10 questions:

	child 1	child 2	child 3	child 4	child 5	child 6	child 7
Age	2;5.26	3;2.1	4;3.18	4;4.8	5;2.13	5;7.18	6;0.13
Q1	○	○	○	○	○	○	○
Q2	○	×	×	○	○	○	○
Q3	○	○	○	○	○	○	○
Q4	×	×	○	○	○	○	○
Q5	×	×	×	×	○	×	○
Q6	×	×	×	○	×	○	○
Q7	×	×	×	○	×	○	○
Q8	×	×	×	○	×	○	○
Q9	×	×	×	×	×	○	○
Q10	×	×	×	○	×	○	○

Key: O = success, X = failure

Table 2: Two, three, and four embedded recursive possessives for Japanese children

Again, the youngest children correctly get a single possessive but then fail. With the exception of example 5 (about shoes, not balls) and child 5 (who manages double but not triple recursive structures), what stands out is that those children who master 3-part possessives (33.7–9) and have no difficulty with 4-part possessives (33.10). The 2-part possessives (33.4–6) are likewise grasped almost at the same time as 3- and 4-part possessives by three quarters of all children, all below 7 years of age, clearly much younger than the L2 college students.

What role do the pictures play? One might observe that they give us a visual hook with which to keep track of all the relations. They are an additional cognitive guide to the meaning. While this is correct, it is not a substitute for grammar, as the child dialogue above, where meaning (whose car) is understood, but recursive expression of it is difficult. If we had the conjoined version,

(34) “Show me Mika’s and brother’s and friend’s and dog’s ball.”

it would call for us to point to all of their balls (which is more work) and not just the final one, just as our first example elicited a reference to Cookie Monster’s *and* sister’s pictures.

This is precisely what transpired with the younger children who failed to grasp the recursive sentences. When there were more than one possessive, child 2’s answers were more than one. For example, for the question, “What color is Sho’s friend’s ball?”, his answer was “This and this and this”, and he pointed to Sho’s ball, Sho’s dog’s ball, and his friend, Yuki’s ball. Other answers among the younger children involved deleting one or more possessive.<sup>21</sup>

<sup>21</sup> In a larger experiment just completed, 26 Japanese children (seven 3-year-olds, eight 4-year-olds, seven 5-year-olds, and four 6-year-olds) were tested by Roeper *et al.* (in progress) with the same basic format. Similar results were obtained with intriguing further detail, but a full analysis must still be done.

In brief, the children were given three pictures and asked 16 questions involving 1- to 4-level recursion (three 1-POSS, five 2-POSS, four 3-POSS, and four 4-POSS). The sentences were linked to pictures where every person or animal had a hat and a balloon of a variety of colors, with questions like:

- |       |  |        |
|-------|--|--------|
| (i)   | What is the color of Shiro’s child’s friend’s dog’s balloon? | 4-POSS |
| (ii)  | What is the color of Murasaki’s friend’s dog’s balloon?      | 3-POSS |
| (iii) | What is the color of Orenji’s dog’s balloon?                 | 2-POSS |

11 children showed mastery (80–100%) of 1- and 2-POSS and showed 50–100% correct at 3- and 4-POSS level recursion, showing that they could in general handle recursion. Interestingly, there was no difference in ability at 3- and 4-level cases. This suggests that an incremental parsing theory cannot account for the difference. Errors often involved deleting one element (which we saw with L2 speakers). This suggests that keeping track of names is an independent challenge beyond recursion, 4 children failed to get 1-level possessives right and 22 got 100% of 1-POSS correct. 4 children got 1-POSS right 100%, but only 2 out of 5 2-POSS cases. This is exactly what the hypothesis that recursion is a separate operation predicts. 7 children get 3 out of 5 2-POSS cases right, but very low success on 3- and 4-POSS. This is not exactly what the hypothesis that recursion allows 2-, 3-, and 4-POSS cases equally would predict. Informal discussion with L2 speakers who find these difficult suggests that they have some ‘strategy’ to compute a second POSS, but they are unable to handle three — indicating that real recursion is not yet in place.

Further observations include that relational nouns, like *friend’s*, might seem to be a point of confusion. We think that part–whole sequences may be where children first have success where phrasal nouns are modified: *the big house’s*, *small porch’s*, *back swing’s color*. The micro-

In sum, we have pointed out evidence that the acquisition of recursion is not immediate, but that once recursion is acquired, there is not a significant processing demand producing a difference between 3- and 4-level possessives.

### 3.4. Verbal Compounds

Snyder (1995) showed that 3- to 4-year-olds produce novel two-word compounds and Hiramatsu *et al.* (2002) showed that the ability was productive. For verbal compounds, Hiraga (2010) found that children at the age of 4 to 5 years were easily able to understand and produce a single verbal compound: When asked “What is someone who pulls a wagon?”, they provided the answer “Wagon-puller”, corroborating claims in Clark (1993). Novel compounds like “I’ll be the lunch-bringer” occur as well at 4 years. When Hiraga sought to see if recursive compounds were possible, much greater difficulty was encountered. Only by the 6- to 7-year-old range did children show clear ability to comprehend. Here is one of the stories and the picture that accompanied it:

- (35) *Kitty makes a great machine. The machine pours tea into many cups at once. Bunny bought the machine from Kitty, so Bunny only makes tea and doesn’t have to pour it. The machine pours tea into five cups at once, so Bunny’s sisters and brothers can drink it. Doggy doesn’t have the machine, so he makes and pours tea himself. One of them said “I am a tea pourer maker”.*

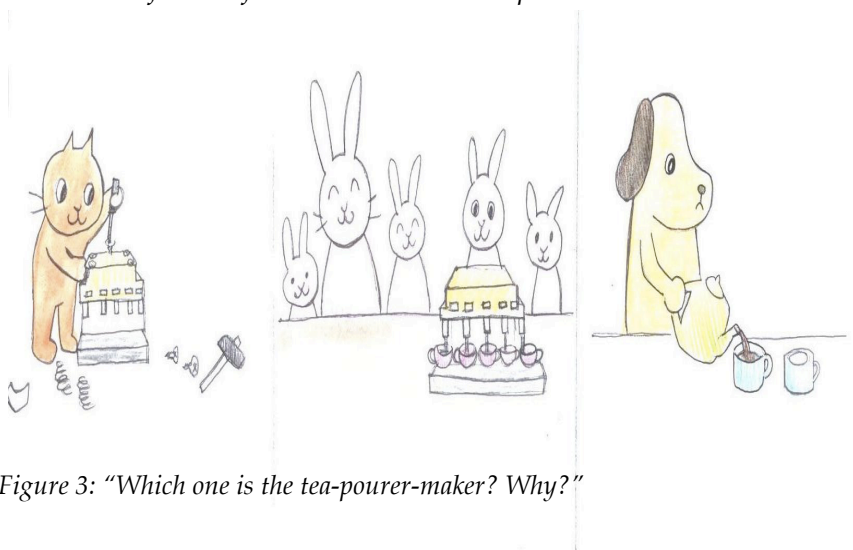


Figure 3: “Which one is the tea-pourer-maker? Why?”

The conjunctive reading is interpreted by 6-year-olds as: “Because he makes and pours tea”. Examples of recursive interpretation follow:

- (36) a. *tea pour maker* (N.I. at 6;11.2):  
 “because she pours, actually, she made the machine that pours tea”  
 b. *tea pourer maker* (N.I. at 6;11.2):  
 “because he makes the machine that pours tea”

---

conditions that provide triggering recursive environments for children are needed to see exactly what moves the child along the acquisition path.

(37) *tea pourer maker* (I.R. at 6;5.4):

“because she made the machine that can pour tea”

(38) *tea pourer maker* (P.H. at 5;11.20):

“because he made the machine that could pour it for you”

Other examples with stories included the following:

(39) *pencil sharpener spiller, picture taker liker, bottle opener breaker, tea lover taker*

These examples feel intuitively more difficult for adults as well, but 9 out of 10 adults gave 89% correct answers; only 2 out of 45 gave a conjunctive reading.<sup>22</sup>

A question arises: Why should this form of recursion be so much later than adjective recursion (*second green ball*)? Why does it feel more difficult to adults? Noun compounds [*school lunch box*] are much more frequent than recursive verbal compounds [*bread-baker watching*]. This means that, although just a few can trigger the process, their rarity could affect when they appear. This does not offer a full explanation of delay however.

We argue elsewhere that the derivational path is relevant: It is a reflection of leftward movement operations and Relativized Minimality (see Friedmann *et al.* 2009). In effect, in *tea-pourer-maker* one compound with *-er* must ‘cross over’ another *-er* [*maker of tea-pourer* → *tea-pourer-maker*]. Cross-over may explain part of why left-branch verbal compounds are especially difficult, but recursion itself seems to be the problem where no cross-over is present in rightward forms of recursion, as in adjectives and PPs and, as we now discuss, complements.

### 3.5. Sentence Complements

Finally, we add sentence complements to our overview, although they engage many more aspects of grammar than simply recursion. The first observation to make is that children appear to acquire infinitives very quickly in a recursive form, although they are arguably not phases, and certainly do not contain propositional content. A cursory search reveals recursive infinitives as early as age 2;4, but a careful study of their emergence would be very useful. Here are examples from a broad search of the database of children below 4;6 (Adam was between 3;6 and 4;6), although exact ages are not obtainable.

(40) Naomi (2;4): *to go to sleep*

Adam 27\*CHI: what you use *to carve* [?] *it to do* what?

Adam30.cha: I want *to go to sleep* and stand up

Adam 54\*CHI: here (.) we *going to have to build* one with another string on it.

Anne 34b\*CHI: you have *to get this one to go* as\_well.

stp2.cha: here am I *going to get to put* the chimney

<sup>22</sup> A kind of conjunctive reading (i.e. *or*) is possible if one takes a ‘slash’ interpretation, as in *This is a printer/copier*. When the experiment was designed, we were unaware of this option, and it is interesting that it was not taken more often by adults.

boys44.cha: when I got bigger then I'm going *to get to go*  
 cha:\*CHI: dad you're suppose *to try to get* it on me  
 e21.cha:\*CHI: Now I'm going *to try to touch* your knee  
 aran 29\*CHI: I went *to climb the house to see* them.  
 \*CHI: I want *to get to see*.  
 nic34b.cha\*CHI: you have *to go to sleep* now.  
 liz22b\*CHI: this one [\* 0is] going *to go down to drink*

How come these forms emerge at such an early point? Do they really represent a series of phases? Evidence of this kind may fit the notion that infinitives are not phases, but much more argument is needed.

Nonetheless when we turn to tensed complements, we both find some at the 4- to 5-years range (from Adam, 4;5 to 4;8) just a few earlier, though a more thorough search would be useful:

- (41) Danilo 3.2 I think Daddy says *he wears underpants*  
 adam45.cha:\*CHI: he thought *those guns were coming from outside him*  
 adam45.cha:\*CHI: he thought *I said something (a)bout... window*  
 adam52.cha:\*CHI: he thought # *bad people would get him*  
 I thought *you said they gonna be warm*

These forms might, however, be represented by a recursive adjunct that is not really sentential, much like:

- (42) to me and for you they gonna be warm

Diessel (2005) argues that the early forms of *I think* simply mean *maybe*. We must establish that each clause is really embedded inside the other, as we do next.

In fact, when we look at real comprehension, we find that children resist complementation. Even when the momentum of the story is clear, the children 'anti-pragmatically' resist embedding. Hollebrandse *et al.* (2008) with 18 children (6;3–6;11, mean: 6;9) have shown that they have no difficulty giving single complement answers for situations with sentences like: "Dad is talking to Billy about moving his tools. Dad tells Billy that Jane said that hammers are too heavy. What did Jane say?" Children easily respond "hammers are too heavy".

However, when the higher verb is needed to make sense of the question, implying recursive subordination, correct answers fall off sharply. Thus among children as old as 6, only one third provide the recursive answer, although the meaning is very misleading if you do not in the following story. Because conjunction can deliver the same inference, we sought each time to find a meaning that guarantees embedded recursive structure:

- (43) *Jane talks to mom. She is having a fight with Billy on the phone. Jane tells mom that Billy said that all sisters are stupid. What did Jane tell mom?*

Confronted with a drawing depicting the setting (Fig. 4), the two possible responses would be the ones given in (44):

- (44) Single complement: [said] “that all sisters are stupid”  
 Recursive complement: [tell Mom] “that Billy said that all sisters are stupid”.

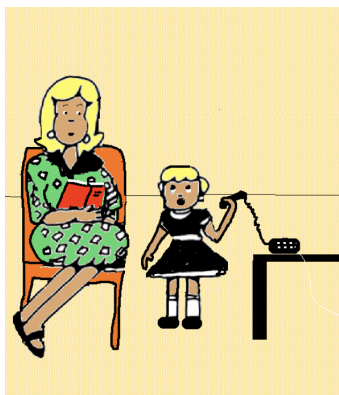


Figure 4: Two complement sentential recursion experimental illustration

The experiment is constructed anti-pragmatically because if she gives a single complement answer (“all sisters are stupid”), then she condemns herself. Here are the results:

- (45) 23% irrelevant, 34% single, 33% recursive

Thus, these children, until the age of 6, despite being invited by the momentum of the story to oppose the boy by mentioning that he is the speaker, offer a single clause or an irrelevant answer in two thirds of the answers. This leads to the clear conclusion that a single complement answer is not represented in the same manner as a recursive complement.

In sum, the children allow a single possessive, single adjective, single complement preferentially as the first step. The second step involves a direct-recursion conjoined response. Finally we obtain an indirect recursive response.

#### 4. Generalized Transformations and Tree-Adjoining Grammar

Now we need to address the question squarely of what change could occur to shift from a conjunctive representation to a recursive one. In principle, recursion is an automatic consequence in a phrase-structure rule system. If one category contains another, then what would block the generation of recursive forms? Thus if I have,

- (46) *John said S2*

and I realize S2 as NP – VP and choose *Bill said* for VP, then I automatically introduce another VP — and *it is raining* is possible, giving:

- (47) *John said Bill said it is raining.*

Under this formalism, we must stop this core process from occurring.

If the initial representation, however, were not an expandable S-node, then the derivation could be constrained to a single complement.<sup>23</sup> In fact, Perfors *et al.* (2011) have suggested that children might begin with a direct subcategorization of complement structure that avoids recursion.<sup>24</sup>

(48) NP *think* NP *verb* NP

This would predict that the child's progress to other forms would occur step by step, just like matrix clauses emerge step-by-step. For instance, the passive form,

(49) *John thinks* NP *was* V+*ed* *by* NP

would have to be separately acquired. Then at some point — a critical point in the biology of the organism — the list of possible structures becomes uneconomical and the child substitutes S or CP for the whole list. At that point recursion would be present and nothing could stop it: *John thinks that Bill thinks that Fred believes...* And each additional instance would not be costly. Capturing that act of substitution is a fundamental acquisition ability. It is not automatically represented in UG.

This substitution approach has independent plausibility when one considers, for instance, how V2 develops (and historical evolution has similarities, see Westergaard 2009). Children begin with locative–verb–subject (*da isst er* 'there eats he'), then other forms, like conjunctions appear (Conj–V–subj) (*nun kann ich* 'now can I'), but only at a late stage does the child acquire full V2, allowing Obj–verb–subj (*Fleisch isst er* 'meat eats he')<sup>25</sup> with XP–verb–subj, where V2 is defined entirely in categorical terms as XP followed by V. In English, V2 exists but in lexical, not categorical terms. Only verbs of speaking are involved: "Nothing"

---

<sup>23</sup> We know that in other environments recursion is blocked, as in evaluatives:

- (i) John knew Bill to be a fool.
- (ii) \* John knew Fred to know Bill to be a fool.

Evaluatives involve personal experience, as in:

- (iii) John knew Fred to be a liar.

which contrast with propositional complements which allow recursion:

- (iv) John knew that Fred was a liar.
- (v) John knew that Bill knew that Fred was a liar.

Therefore, the grammar must have a method to block recursion for evaluative complements. This happens automatically if we turn the logic around: the grammar should not allow recursion unless there is an explicit example.

<sup>24</sup> Tenenbaum and colleagues see this form of incrementalism as linked to data-processing procedures derived from general learning psychological theories. The critical moment of substitution suggests the opposite from my perspective: the move to a higher order category like Sentence, or CP, or any category that covers a superficially heterogeneous set of strings into a single higher category node label only succeeds by positing an organism with an innate bias toward specific abstract categories like CP. It would be a mistake to put all phenomena into one generalization: *think* [X]. Then: *He thinks quickly* and *He thinks he is sleeping* would lead to a generalization that had *quickly* and *He is sleeping* as one (sentence?) category.

<sup>25</sup> See Roeper (1999), Yang (2002); also Wexler (to appear), de Villiers & Roeper (to appear).

*screamed Bill*. Notably, it does not allow recursion: \*"*nothing*" *screamed Bill*" *screamed John*. Non-recursion follows if we generalize a local notion of subcategorization, allowing lexically specific or low-level categories to be locally generated. This is what allows idioms and very limited clausal subcategorization to exist in adult language:

- (50) a. you were supposed to do that  
 b. \*I supposed you to do that.

where *suppose* only allows the passivized form to take a complement. Such lexically restrictions never extend into recursive domains and no category exists that could expand into recursion.

TAG develops a notion of substitution where a non-recursive node is substituted for by a recursive node. However, this operation of substitution in acquisition is not identical to those in TAG because the criterion of Label Identity is not met: "[T]he substitution operation imposes a requirement of label identity between the root of the substituted elementary tree and the substitution site" (Frank 2006: 149).

It is, though, a fundamental aspect of microscopic steps in acquisition growth. Therefore we have to have a more powerful method for the acquisition device to establish equivalence between a string and a higher category:

(51) *Acquisition Substitution Algorithm*

Substitute a UG Category for a set of strings

This is where innate UG assumptions are needed to make acquisition efficient. We need a substantive notion of S(entence) allowing the projection of a higher category from a set of possible strings with T(ense) and VP at the core: NP T VP → S. This question is really the acquisition version of how we develop a Labeling Algorithm, which Chomsky (2008) has proposed, but which remains largely unarticulated. This challenge reaches to the heart of linguistic theory because the system must not allow a nominalization, for instance, to be analyzed as a sentence. If it did, then the child would generalize *John eats Bill's cooking* into *eat* (S) which would incorrectly allow the generation of \**John eats Bill cooks*.

TAG proposes a more general form of substitution that may be useful for other cases:

I will adopt the two operations of the Tree Adjoining Grammar (TAG) formalism: *substitution* and *adjoining* [...]. Given two independently derived pieces of structure, the substitution operation inserts one along the periphery (or frontier) of the other at a node, called the substitution site. One can think of substitution as an operation which rewrites a node along the frontier of one structure as another piece of structure (called an *elementary tree*).

(Frank 2006: 149)

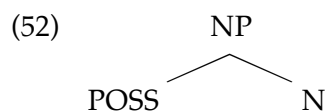
This approach obeys a principle of Label Identity:

I assume the Condition on Elementary Tree Minimality (CETM), according to which the heads in an elementary tree must form part of the extended projection of a single lexical head, following the notion of extended pro-

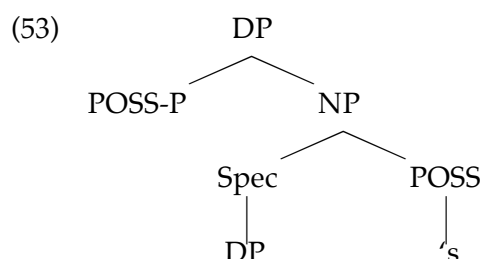


jection of Grimshaw (2000). As a result, the lexical array underlying an elementary tree must include a lexical head and may include any number of functional heads that are associated with that lexical head. (Frank 2006: 151)

We suggest that a simple form of projecting a Maximal Projection instead of a single lexical category is a natural projection of this substitution:



and we now make POSS into an MP with a SPEC position:



And this will automatically allow in principle a recursive projection. If [Spec, POSS-P] carries a feature that allows D projection, a DP can be projected and recursion is launched. Thus a specific operation would be present to accomplish this goal.

We argue in effect that the addition of new projections both allows recursion and reflects a distinct computational act along the acquisition path. Once introduced into the grammar, however, each further instance will be automatic and therefore we do not predict that the shift from two to three to four embeddings will cause a serious online increase in difficulty. Under classic Generalized Transformations, each time a complex POSS phrase is added, a substitution must occur, which if translated into a processing account, would predict incremental difficulties. Longer sentences always produce some parsing complexity, but our evidence suggests that additional possessives do not complicate matters. Therefore, in this case, the acquisition substitution into the equivalent of rewriting rules is sufficient.

#### 4.1. *Relative Clause Substitution*

Frank's account of relative clause attachment is similar:

- (54)  $S \rightarrow DP_1 VP \Rightarrow$  John like  
 $VP \rightarrow V DP_2 \Rightarrow$  John liked the cat.

However, independent of this form we have a second rule:

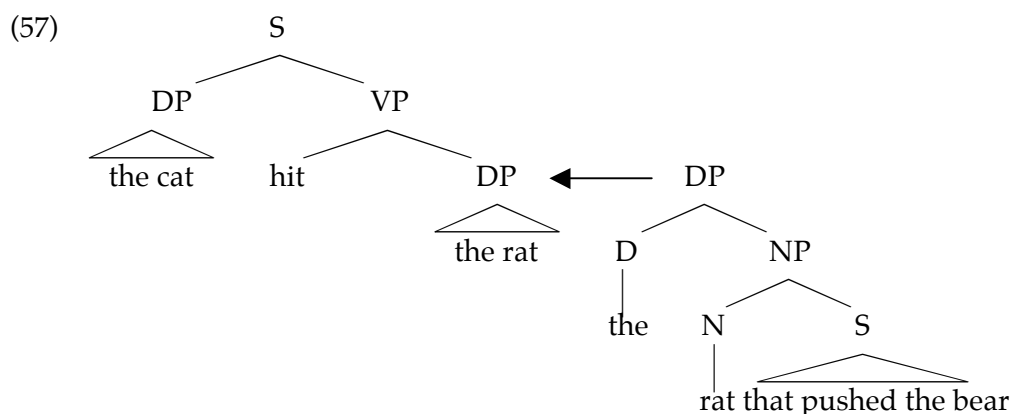
- (55)  $DP_3 \rightarrow NP S$

which carries a branching node and a meaning that allows the relative clause to

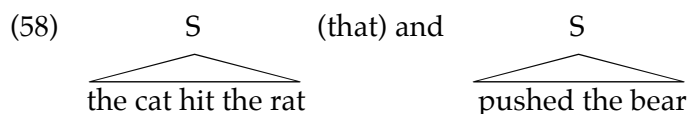
restrict the range of reference, therefore to participate in the interpretation of the DP. Thus TAG allows the generation of two forms:

- (56) Sentence: DP1 [the rat] VP [hit] DP2 [the cat]  
 DP3 [the cat] Sentence [that I like]

and the second tree is inserted by substitution of DP3 for DP2 into the first:



Without the substitution, the relative clause automatically attaches as an adjunct to the Root node and we have exactly conjunction as we found in the examples above and as Goodluck (1981) argued, who said *that* was treated as *and*:



Predictably, as the earliest results showed (Tavakolian 1981), the relative clause is typically interpreted with reference to the subject *the cat* instead of *the rat* by children in the 3- to 4-year-old range.<sup>26</sup>

#### 4.2. Labeling Algorithm

The notion of substitution of a complex form interacts with the current lively question of how labels are determined. If our proposal is carried forth, it will require that the Labeling Algorithm be one that fits this move (see Chomsky 2008). Capturing the acquisition path might in fact be an important criterion to evaluate the formulation of a labeling algorithm. In effect it would be a method whereby recursive nodes could look different from non-recursive ones which in turn would fit the claim we make that the acquisition path for recursion involves a critical step beyond recognition of the basic syntactic category. The acquisition path should reflect upon the notational choices made in linguistic theory.

<sup>26</sup> Many grammars (Keenan & Comrie 1977) allow a final relative clause, attached to the root, to be interpreted either with the subject or the object.

## 5. The Experience of Recursion

We are now in a position to answer the question raised by William Snyder and me in a series of papers examining the appearance in naturalistic data of recursive compounds, possessives, adjectives, and serial verbs. We advanced the hypothesis that children must ‘experience’ recursion in order to allow it in their language (Roeper & Snyder 2004, 2005). We had no statement about what impact the experience of recursion would cause.

This hypothesis followed from the observations above, that single instances of possessives, adjectives, and compounds in a language did not guarantee recursion.<sup>27</sup> If we now argue that recursive nodes are discernibly different from non-recursive nodes, then the argument that experience is necessary is clearly justified. A consequence, of course, is that such triggers are rare, and hence we can predict that they may arise late or in a non-uniform fashion among children. The number of times one hears *coffee-maker* in comparison to *coffee-maker-maker* is obviously small. If recursion is the primary form of productivity in grammar, the rare evidence for recursion becomes a powerful demonstration that frequency is not a primary factor in advancing productive powers in grammar. In fact, children occasionally spontaneously create recursion in new environments, but not often, suggesting that the experience requirement is correct.<sup>28</sup>

This leads to the question whether language-specific recursion is a marginal phenomenon — as much of the public controversy would suggest (Everett, to appear) — or whether recursion is the fundamental pivot, the axis which forces productivity and allows an efficient flow of thoughts into language.

It is not simply an abstract question. A close look supports the latter view. Recursive operations operate upon hierarchical structures. Those labeled hierarchical structures represent a range of abstractions that allow some productivity. Identifying a node with an NP allows any NP to occupy that node. As we have seen, subcategorization (which applies to verbs but also to other lexical items) allows the hierarchy to be overruled by lexically specific information. Thus, the verb *crane* allows only *necks* as an object; you cannot *\*crane your elbow*. Recursion, once recognized by the child, never allows this constraint: It operates only on grammatical categories. A single complement may be an idiom: *John knows what’s what*. It is not possible for *know* to project such an idiom into a recursive domain, that is, over another clause: *\*John knows Bill thought what’s what*.

The recognition of recursion is an automatic liberation from searching for idiomatic subcategorizations. And it relegates exceptional constructions to secondary grammars. Under the Multiple Grammars approach (see Roeper 1999), a signal feature of the presence of a sub-grammar is the absence of recursion. An example again is V2 in English, discussed above, which applies to quotations (“*Nothing*” said Bill) and stylistic inversion (*In the room ran John*) — but notably neither allows recursion.

This bifurcation between recursive and non-recursive rules gives the child

<sup>27</sup> The idea originated with the observation about productivity from Namiki (1994) that only those grammars with recursive compounds had productive compounds.

<sup>28</sup> For example, a 4-year-old said *Here is another another box*, where adults would say ‘yet another box’. See my *Prism of Grammar* (Roeper 2007) for a few other examples.

a means to assemble his core grammar and exclude marginal exceptions. Before specific nodes, entailing recursion are recognized, it is commonly suggested (Roeper 1992, Tomasello 2003) that there is a great deal of lexical specificity that blocks or limits easy overgeneralizations. To return to our possessive example, the child may first represent possessives in both English and German with a constraint on human or animate possession. Independently assembling examples of early possessives, from Galasso (1999), here is what we find:

- (59) a. i. I want *me* bottle. Where *me* Q-car? That *me* car. Have *me* show. *Me* turn. *Me* cat. *Me* pen. (2;6–2;8)  
 ii. No *you* train. It's *you* pen. It's *you* kite. It *you* house? (3;2)  
 iii. I want to go in *him* house. *Him* bike is broken. It's *him* house.  
 b. *Lexical*  
 i. *Mine* banana. *Mine* bottle. *Mine* car. *Mine* apple. *Mine* pasta (2;4)  
 ii. *My* car. (3x at 2;4) *My* pasta. I want *my* key. It is *my* t.v.

(60) *Single Poss*

[whose hat is that] "Mrs. Wood's" (2.7)

Jensen & Thornton (2007)

They are all human possessors (no cases like *the car's tire*) and, of course, none are recursive. Therefore at Stage 1, the English and German child may have the same grammar. When the child re-analyzes the possessive to allow recursion, as we saw above, then the grammars diverge, and lexical constraints on the nature of the NP are dropped. Thus we suggest that it is exactly recursion which enables the child to generate a grammar where English and German diverge.

## 6. Conclusion

Our focus has been various language-particular forms of recursion. We have seen a variety of evidence of a default conjunctive interpretation that can be captured by Direct Recursion: the included possessives, PPs, adjectives, and complements.

We claimed that the core analysis lies in a combination of Indirect Recursion and SMT, the Strong Minimalist Thesis. Finally, we sought to explain why recursion is not immediate via the proposal that Generalized Transformations cause the definition of recursive nodes to be distinct from non-recursive ones such that an operation of Substitution is necessary, as proposed in TAG:

Implicit in the study are several broader claims:

- A. If variation exists in where languages allow recursion, then an acquisition challenge exists.  
 B. The grammar, not simply processing, can be engaged in formally specific ways to capture this acquisition path which, moreover, provide insights into the formalisms themselves.

- C. The time-course of each form of recursion may be a function of how much exposure is involved, the nature of the derivation, the intersection with morphology, and other factors.
- D. The representation of recursion critically involves an interface with interpretation — via phases and the SMT — which we take to be an innate interface.
- E. Our mode of argumentation, given the obscurity of the process and the evidence, is to include small amounts of suggestive evidence if they point in the same direction and contribute to a deeper generalization, or acquisition hypothesis. This then invites a more thorough program of research.

In sum, we argue that the child seeks many kinds of recursion as the core of syntactic productivity.

## References

- Berwick, Robert C. & Noam Chomsky. 2011. The Biolinguistic Program: The current state of its evolution and development. In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, 19–41. Oxford: Oxford University Press.
- Berwick, Robert C. & Amy Weinberg. 1986. *The Grammatical Basis of Linguistic Performance*, Cambridge, MA: MIT Press.
- Boeckx, Cedric. 2009. On the locus of asymmetry in UG. *Catalan Journal of Linguistics* 8, 41–53.
- Bryant, Doreen. 2006. *Koordinationsellipsen im Spracherwerb: Die Verarbeitung potentieller Gapping-Strukturen*. Berlin: Akademie-Verlag.
- Chomsky, Noam. 1965. *Aspects of a Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria L. Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2010. Fall lectures. Cambridge, MA: Massachusetts Institute of Technology.
- Clark, Eve V. 1993. *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.
- Diessel, Holger. 2004. *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press.
- Everett, Daniel L. To appear. You drink. You drive. You go to jail. Where's recursion? In Margaret Speas & Tom Roeper (eds.), *Recursion*. New York: Oxford University Press [under review].
- Frank, Robert. 2006. Phase theory and Tree Adjoining Grammar. *Lingua* 116, 145–202.

- Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88.
- Fujimori, C. 2010. Acquisition of recursive possessives in Japanese. Ms., University of Massachusetts, Amherst.
- Galasso, Joseph. 1999. The acquisition of functional categories: A case study. Colchester: University of Essex dissertation.
- Gentile, Sarah. 2003. On the acquisition of left-branching recursive possessives. Amherst, MA: University of Massachusetts honors thesis.
- Goodluck, Helen. 1981. Children's grammar of complement-subject interpretation. In Susan L. Tavakolian (ed.), *Language Acquisition and Linguistic Theory*, 139–166. Cambridge, MA: MIT Press.
- Grimshaw, Jane. 2000. Extended projection and locality. In Peter Coopmans, Martin Everaert & Jane Grimshaw (eds.), *Lexical Specification and Insertion*, 115–133. Amsterdam: Benjamins.
- Gu, Chenjie. 2008. Structural ambiguity and AP/PP recursion in language acquisition. Ms., University of Massachusetts, Amherst.
- 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.
- Hiraga, Misato. 2010. Acquisition of recursive verbal compound nouns. Paper presented at the 32<sup>nd</sup> Annual Conference of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS), Humboldt University, Berlin (24 February 2010).
- Hiramatsu, Kazuko, William Snyder & Tom Roeper. 2002. Learnability in a Minimalist framework: Root compounds, merger, and the syntax–morphology interface. In Ingeborg Lasser (ed.), *The Process of Language Acquisition*, 25–37. Frankfurt: Peter Lang.
- Hollebrandse, Bart, Kathryn Hobbs, Jill G. de Villiers & Tom Roeper. 2008. Second order embedding and second order false belief. In Anna Gavarró & M. João Freitas (eds.), *Proceedings of Generative Approaches to Language Acquisition (GALA) 2007*, 270–280. Newcastle: Cambridge Scholars Publishing.
- Hornstein, Norbert. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.
- van Hout, Angeliek, Masaaki Kamiya & Tom Roeper. 2011. Passive, reconstruction and edge phenomena: Connecting English and Japanese nominalizations. Ms., Rijksuniversiteit Groningen, University of Massachusetts, Amherst & Hamilton College, Clinton, NY.
- Jackendoff, Ray. 2010. What's in the human language faculty: Two views. Ms., Tufts University, Boston, MA.
- Jensen, Britta & Rosalind Thornton. 2007. Fragments of child grammar. Ms., Macquarie University, Sydney.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63–99.
- Limbach, Maxi & Dany Adone. 2010. Language acquisition of recursive possessives in English. *Boston University Conference on Child Language (BUCLD)* 34, 281–290.

- Lobina, David J. 2010. What recursion could not be. Ms., Universitat Rovira i Virgili, Tarragona.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk* (3<sup>rd</sup> edn.). Mahwah, NJ: Lawrence Erlbaum.
- Namiki, Takayasu. 1994. Subheads of compounds. In Shuji Chiba (ed.), *Synchronic and Diachronic Approaches to Language: A Festschrift for Toshio Nakao on the Occasion of His Sixtieth Birthday*, 269–285. Tokyo: Liber Press.
- Matthei, Edward H. 1982. The acquisition of prenominal modifier sequences. *Cognition* 11, 301–332.
- Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, 419–491. New York: Wiley.
- Munn, Alan. 1999. First conjunct agreement: Against a clausal analysis. *Linguistic Inquiry* 30, 643–668.
- Perfors, Amy, Joshua B. Tenenbaum & Terry Regier. 2011. The learnability of abstract syntactic principles. *Cognition* 118, 306–338.
- Pietroski, Paul M. 2011. Minimal semantic instructions. In Cedric Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, 472–498. Oxford: Oxford University Press.
- Richards, Marc D. 2011. Deriving the edge: What's in a phase? *Syntax* 14, 74–95.
- Roeper, Tom. 1978. Linguistic universals and the acquisition of gerunds. *University of Massachusetts Occasional Papers in Linguistics* 4, 1–37.
- Roeper, Tom. 1992. From the initial state to V-2: The architecture of acquisition. In Jürgen M. Meisel (ed.), *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*, 333–370. Dordrecht: Kluwer.
- Roeper, Tom. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2, 169–186.
- Roeper, Tom. 2007. *The Prism of Grammar: How Child Language Illuminates Humanism*. Cambridge, MA: MIT Press.
- Roeper, Tom. 2009. The minimalist microscope: How and where interface principles guide acquisition. *Boston University Conference on Child Language (BUCLD)* 33, 24–48.
- Roeper, Tom & Ana T. Pérez-Leroux. 2010. Simplicity and complexity in child language and its explanation. Ms., University of Massachusetts, Amherst & University of Toronto.
- Roeper, Tom & William Snyder. 2004. Recursion as an analytic device in acquisition. In Jacqueline van Kampen & Sergio Baauw (eds.), *Proceedings of Generative Approaches to Language Acquisition (GALA) 2003*, 401–408. Utrecht: LOT.
- Roeper, Tom & William Snyder. 2005. Language learnability and the forms of recursion. In Anna Maria Di Sciullo (ed.), *UG and External Systems: Language, Brain and Computation*, 155–169. Amsterdam: John Benjamins.
- Roeper, Tom, Akiko Terunuma & Terue Nakato-Mishiyata. In progress. Acquisition of recursive possessives in Japanese. Ms. University of Massachusetts, Amherst.
- Ross, John R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT

- dissertation.
- Sauerland, Uli. 2010. Experimental evidence for complex syntax in Pirahã. Ms., Zentrum für allgemeine Sprachwissenschaft, Typologie und Universalienforschung, Berlin. [<http://ling.auf.net/lingBuzz/001095>]
- Snyder, William & Tom Roeper. 2004. Learnability and recursion across categories. *Boston University Conference on Child Language (BUCLD) 28*, 543–552.
- Stabler, Edward P. To appear. Recursion in grammar and performance. In Margaret Speas & Tom Roeper (eds.), *Recursion*. New York: Oxford University Press [under review].
- Tavakolian, Susan. 1981. The conjoined-clause analysis of relative clauses. In Susan Tavakolian (ed.), *Language Acquisition and Linguistic Theory*, 167–187. Cambridge, MA: MIT Press.
- Tomalin, Marcus. 2007. Reconsidering recursion in syntactic theory. *Lingua* 117, 1784–1800.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard: Harvard University Press.
- Villiers, Jill G. de and Tom Roeper. In press. Introduction. In Jill G. de Villiers & Tom Roeper (eds.), *Handbook of Generative Approaches to Language Acquisition*. New York: Springer.
- Westergaard, Marit. 2009. *The Acquisition of Word Order: Input Cues, Information Structure and Economy*. Amsterdam: John Benjamins.
- Wexler, Kenneth. In press. Grammatical computation in the optional infinitive stage. In Jill de Villiers & Tom Roeper (eds.), *Handbook of Generative Approaches to Language Acquisition*. New York: Springer.
- Yang, Charles D. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Tom Roeper  
University of Massachusetts  
Department of Linguistics  
South College 218  
Amherst, MA 01003  
USA  
[roeper@linguist.umass.edu](mailto:roeper@linguist.umass.edu)



# The Neural Basis of Recursion and Complex Syntactic Hierarchy

Angela D. Friederici, Jörg Bahlmann,  
Roland Friedrich & Michiru Makuuchi

Language is a faculty specific to humans. It is characterized by hierarchical, recursive structures. The processing of hierarchically complex sentences is known to recruit Broca's area. Comparisons across brain imaging studies investigating similar hierarchical structures in different domains revealed that complex hierarchical structures that mimic those of natural languages mainly activate Broca's area, that is, left Brodmann area (BA) 44/45, whereas hierarchically structured mathematical formulae, moreover, strongly recruit more anteriorly located region BA 47. The present results call for a model of the prefrontal cortex assuming two systems of processing complex hierarchy: one system determined by cognitive control for which the posterior-to-anterior gradient applies active in the case of processing hierarchically structured mathematical formulae, and one system which is confined to the posterior parts of the prefrontal cortex processing complex syntactic hierarchies in language efficiently.

*Keywords:* Broca's area; hierarchy; recursion

## 1. Introduction

In the long-standing discussion of what it means to be human, language has always been considered a major component. Recently, the debate has clustered around the question to what extent recursion can be considered as the crucial part of language distinguishing human language from other communicative systems (Hauser *et al.* 2002, Jackendoff & Pinker 2005).

In the context of this discussion, a number of empirical studies on grammar processing have been conducted both in humans and non-human animals. A number of these have used very similar grammar types inviting a comparison between the different animals and cognitive domains. One of the studies directly compared grammar learning in humans and non-human primates, that is, cotton-top tamarins, and reported that non-human primates can learn a simple probabilistic grammar  $(AB)^n$ , called Finite State Grammar, but not a more complex gram-

---

The research reported here was supported by The German Ministry of Education and Research (BMBF) (Grant Nr. 01GW0773).



mar  $A^nB^n$ , called Phrase Structure Grammar.

The recursive structure  $A^nB^n$  is derived from the two rewriting rules below.

- (1)
  - a. Rule 1  $S \rightarrow AB$
  - b. Rule 2  $S \rightarrow ASB$  (a rule for recursion),  
where  $S$  is a non-terminal symbol, and  $A$  and  $B$  are terminal symbols.

$A^nB^n$  is derived for example as in (2):

- (2)  $S \rightarrow$  (with rule 1)  $ASB \rightarrow$  (with rule 2)  $AASBB \rightarrow \dots$  (repeating the rule 2)  $\dots$   
 $\rightarrow A^{n-1}SB^{n-1} \rightarrow$  (with rule 1)  $A^nB^n$

Humans instead easily learned both types of grammar after short training periods (Fitch & Hauser 2004). Interestingly, songbirds were also shown to be able to learn both grammar types, but only after extensive training (Gentner *et al.* 2006). This finding suggests that different species may use different brain systems to solve the same task. The  $A^nB^n$  grammar used in these two studies was not declared to be a test for recursion, but it has been taken to be so by some scientists (Perruchet & Rey 2005, Gentner *et al.* 2006). A recent paper, tries to clarify this issue by defining the term 'recursion' as a rule "which has the property of self-embedding, that is, in which the same phrase type appears on both sides of a phrase structure rewrite rule" (Fitch 2010: 78)

When considering the biological basis of recursion, one has to take this definition into account. Thus it appears that whether an  $A^nB^n$  grammar is recursive depends on the underlying structure. An  $A^nB^n$  grammar could be described as recursive, but does not have to. Fitch (2010) discusses that in the latter case, the assumed processing mechanism, however, must go beyond a finite-state grammar process as it requires "some additional memory mechanism(s) to keep track of 'n'" (p. 87). We will keep this in mind when reporting some recent neuroimaging studies in humans which have tried to evaluate the neural basis of processing different types of grammar, including embedded structures which unambiguously qualify as a test for recursion. These studies used similar syntactic structures in artificial grammar, natural language and non-language domains.

## 2. Finite-State vs. Phrase Structure Grammar

In the first neuroimaging experiment referred to here (Friederici *et al.* 2006a), we investigated the neural basis of grammar processing in humans for the two types of grammar originally used in the behavioural study by Fitch & Hauser (2004) with human and non-human primates, namely an  $A^nB^n$  and an  $(AB)^n$  grammar (see Figure 1).

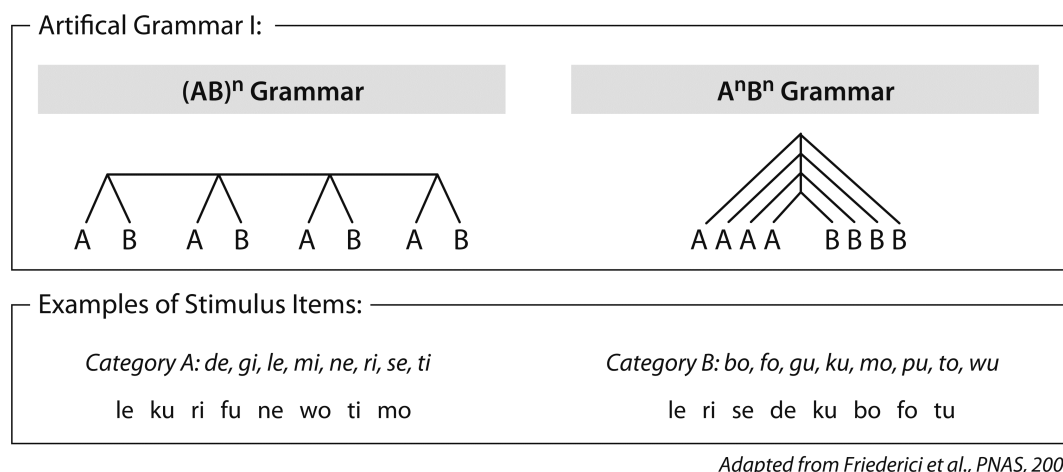


Figure 1: Processing hierarchy in Artificial Grammar I. Structure of sequences is given in the upper row. Category A syllables and Category B syllables used in the sequences as well as examples of an (AB)<sup>n</sup> sequence (left) and an A<sup>n</sup>B<sup>n</sup> sequence (right) are given in the lower row.

In this functional magnetic resonance imaging (fMRI) experiment, category membership was coded by a particular combination of consonants and vowels, and not by pitch information as it was done in the original experiment. Stimulus sequences were presented visually syllable-by-syllable (for details see Friederici *et al.* 2006a). The two grammars were learned by different groups of participants to prevent possible confusion between the two grammars in the participants. During learning, feedback was given. Learning took place two days before scanning.

In the scanning session, grammatically correct and incorrect sequences were presented. The two grammar types led to different activation patterns. The comparison of incorrect versus correct sequences led to activation in the frontal operculum for the (AB)<sup>n</sup> grammar, whereas the comparison of incorrect versus correct sequences for the A<sup>n</sup>B<sup>n</sup> grammar revealed activation in Broca's area (BA 44) in addition to activation in the frontal operculum. This difference was considered interesting in its own right, but, moreover, to be of special phylogenetic importance, since the frontal operculum is considered a phylogenetically older cortex than the more laterally located Broca's area (Sanides 1962).

Thus, it appears that the processing of the more complex artificial grammar with the A<sup>n</sup>B<sup>n</sup> structure recruits the phylogenetically younger cortex, namely Broca's area stronger than the processing of the less complex grammar. Broca's area is known to support syntactic processes in natural language comprehension as evidenced in several studies across different languages (for reviews, see Friederici 2004, Grodzinsky & Friederici 2006, Vigneau *et al.* 2006). The sentences used in the different studies reviewed in these articles include a broad variety of complex syntactic structures such as cleft sentences, passive sentences, scrambled sentences and others, thereby suggesting that Broca's area is involved in the processing of complex hierarchically structured sequences.

From the data reported in Friederici *et al.* (2006a), however, it is not clear whether participants in this experiment did reconstruct a hierarchical embedded structure while processing the A<sup>n</sup>B<sup>n</sup> sequences, or whether the A<sup>n</sup>B<sup>n</sup> sequences were processed by a simple counting mechanism. For example, counting the

number of A elements which then have to be followed by the same number of B elements. Such a mechanism has been claimed to account for the successful processing of the sequences used by Fitch & Hauser (2004) and by Friederici *et al.* (2006a) (see Perruchet & Rey 2005, de Vries *et al.* 2008).

This point is well taken, but given the available literature on syntactic processing which systematically shows an involvement of Broca's area, the observed activation in Broca's area in the present fMRI experiment may suggest that participants did build a hierarchical structure on the basis of which the violation was detected. But this had to be shown in an additional experiment. Moreover, it had to be considered, that the activation in Broca's area could be due to memory processes which are more demanding for the processing of  $A^nB^n$  sequences than for the  $(AB)^n$  sequences used in this study, since the A and B elements were always adjacent in the latter sequences, but not in the former. These open issues were addressed in two subsequent experiments.

### 3. Processing Syntactic Hierarchy

In order to answer the question about the nature of the underlying processes when dealing with  $(AB)^n$  structures, a second fMRI experiment (Bahlmann *et al.* 2008) was conducted in which the sequences were build such that hierarchical processing for the  $A^nB^n$  structures was induced, e.g.,  $[A_1[A_2[A_3 B_3]B_2]B_1]$ . Each subcategory (e.g.,  $A_1$ ,  $A_2$ , etc.) had more than one member to prevent item-based learning. The crucial relations between the dependent elements in the structure were coded by phonological parameters of the respective syllables (see Figure 2).

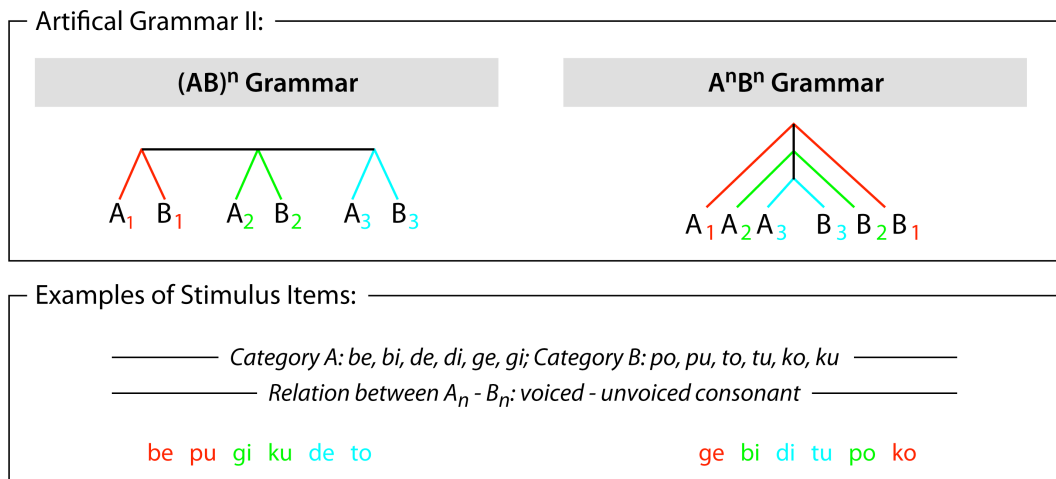


Figure 2: Processing hierarchy in Artificial Grammar II. Structure of sequences is given in the upper row. Category A syllables and Category B syllables used in the sequences as well as examples of an  $(AB)^n$  sequence (left) and an  $A^nB^n$  sequence (right) are given in the lower row. Each subcategory (i.e.  $A_1$ ,  $A_2$ , etc.) comprised two syllables. Note that the relation between  $A^n$ - $B^n$  is defined by the voice-unvoiced dimension of the consonant of the respective syllable.

In this experiment, both grammar types were learned by the same participants to allow a direct comparison of the two grammar types in a within-subject design. This also enabled us to conduct analyses for the correct sequences only in order to evaluate to what extent the observed activations are triggered by grammar processing rather than by the detection grammatical incorrectness.

The direct comparison of brain activation for the two grammar types indicated activation of Broca's area (BA 44), both when collapsed over incorrect and correct sequences, and also when comparing only the correct sequences of the two grammar types (see Table 1 and Figure 4, below). This finding was taken to indicate that the processing of complex hierarchical structures in an artificial grammar involves Broca's area. The result provides support for the interpretation that the processing of the  $A^nB^n$  structures in the experiment by Friederici *et al.* (2006a) reported above was based on hierarchy building rather than on counting plus memory processes needed to keep track of 'n'.

#### 4. Syntactic Hierarchy and Working Memory

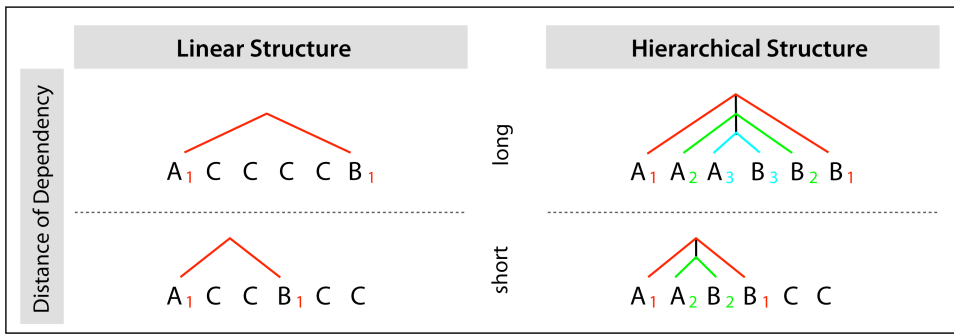
As a second open issue in the interpretation of our initial results, was the question to what extent the observed brain activation was due to working memory involved in the processing of embedded structures, rather than to the syntactic structures as such.

This question is of particular relevance since verbal working memory is known to activate the ventrolateral prefrontal cortex including Broca's area (Jonides *et al.* 1998, Smith & Jonides 1998, 1999), and since it has been claimed that working memory and syntax interact in Broca's area when syntactically complex sentences are processed (Cooke *et al.* 2001, Santi & Grodzinsky 2007). And indeed working memory needs to be considered, as the  $(AB)^n$  and the  $A^nB^n$  structure sequences tested here not only differ in their underlying structure, but moreover in the distance between the dependent A-elements and B-elements. In the studies reported so far the  $(AB)^n$  structure, the distance was always short, since A and B are adjacent, whereas this was not the case for the  $A^nB^n$  structure sequences. Thus, the issue of a possible involvement of memory processes is still unresolved by the prior experiments.

In a further fMRI study (Makuuchi *et al.* 2009, Friederici *et al.* 2009), we investigated to what extent activation in Broca's area is a response to processes of syntactic hierarchy or to working memory. Moreover, we wanted to see to what extent the brain activation pattern observed for artificial grammar processing generalizes to natural language.

The study used German as the testing ground as it allows the construction of sentences with multiple embeddings similar to the previous artificial grammar experiment, e.g.,  $[A_1[A_2[A_3 B_3]B_2]B_1]$  in the form of subject-verb dependencies, e.g.,  $[S_1[S_2[S_3 V_3]V_2]V_1]$  (Figure 3). In order to disentangle the possible confound of the factor syntactic hierarchy and the working memory resources required when dealing with long distance dependencies (e.g.,  $A_3-B_3$ ), we designed a sentence reading study in a 2x2 factorial design, with the factors *syntactic hierarchy* (number of embeddings) and *verbal working memory* (distance of dependent elements).

(a) Experimental design & Examples of Stimulus Items



Peter wusste, dass ...  
*Peter knew that ...*

— hierarchy & long distance —  
[Maria, [die Hans, [der gut aussah], liebte], Johann geküsst hatte.]  
*Maria who loved Hans who was good looking kissed Johann.*

— hierarchy & short distance —  
[Maria, [die weinte], Johann geküsst hatte] und zwar gestern Abend.]  
*Maria who cried kissed Johann and that was yesterday night.*

— linear & long distance —  
[Achim den großen Mann gestern am späten Abend gesehen hatte.]  
*Achim saw the tall man yesterday late at night.*

— linear & short distance —  
[Achim den großen Mann gesehen hatte und zwar am Abend.]  
*Achim saw the tall man at night and that was late.*

Adapted from Makuuchi et al., PNAS, 2009

(b) Tree structure of Stimulus Item (hierarchy & long distance)

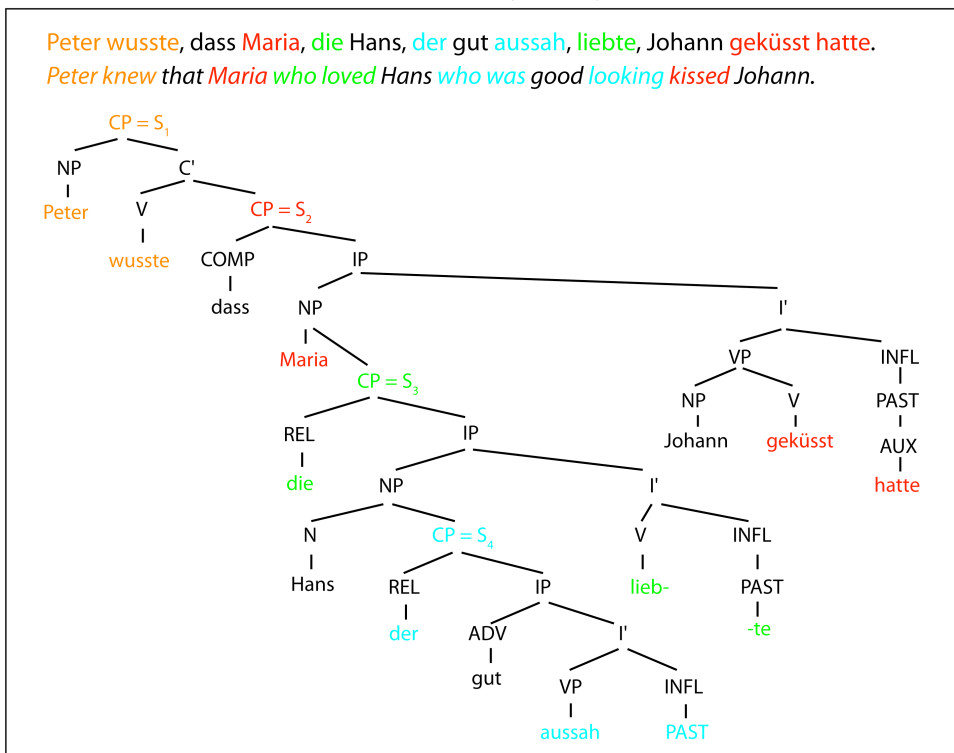


Figure 3:

(a) Processing hierarchy in Natural Grammar. Top: Schematic view of the different conditions. Bottom: Examples of stimulus items for each condition and schematic view of relation between subjects (S) and verbs (V) of (embedded) sentences. Dependent items are color-coded (red, green, blue).

(b) The linguistic description of a sentence used in the natural grammar study (Makuuchi et al., in press). This sentence represents the most complex condition (Hierarchical Structure, long-distance dependency; compare Figure 2).

Key: ADV = adverb, AUX = auxiliary, C = clause, COMP = complementizer, INFL = inflection, IP = inflectional phrase, N = noun, NP = noun phrase, PAST = past tense, REL = relative pronoun, S = sentence, V = verb, VP = verb phrase.

Syntactic hierarchy, as defined by the number of embeddings, activated Broca's area in the inferior frontal gyrus (IFG). In addition, the left superior temporal gyrus (STG) and the superior temporal sulci (STS) indicating that these regions are part of the language network (Friederici et al. 2009). A region of interest analysis of the IFG (Makuuchi et al. 2009) revealed that the main effect of hierarchy was located in BA 44 as defined cytoarchitectonically according to Amunts et al. (1999). In contrast, working memory operationalized by the factor distance between the dependent elements activated the left inferior frontal sulcus located dorsally to Broca's area (see Table 1 and Figure 4). A functional connectivity analysis revealed that these two areas strongly interact during processing multiple embedded sentences.

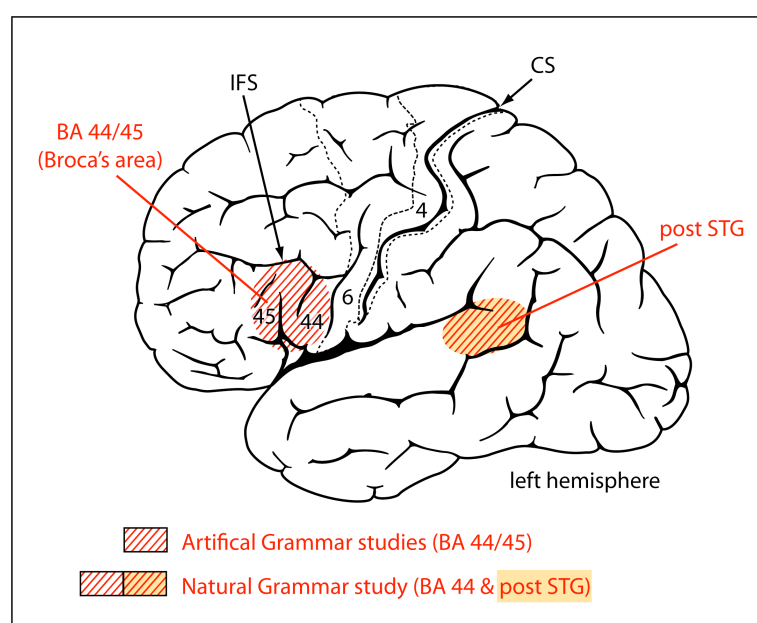


Figure 4: Schematic view of activation pattern for the main effect of hierarchy in the language domain. For Artificial Grammar I and II, the main effect of hierarchy was found in Broca's area (BA 44/45) (Friederici et al. 2006, Bahlmann et al. 2008). For the natural grammar, the main effect of hierarchy was located in BA 44 (Makuuchi et al. 2009) and in the posterior superior temporal gyrus (pSTG) extending into the superior temporal sulcus (Friederici et al. 2009).

Key: BA = Brodmann Area; CS = central sulcus; IFS = inferior frontal sulcus; preSMA = pre-supplementary motor area; STG = superior temporal gyrus.

This locus of activation in Broca's area for the embedded structures coincides with the view that Broca's area supports the processing of syntax in general (Grodzinsky & Friederici 2006). Most recently, a subdivision of syntactic computations within Broca's area for complex syntactic structures has been demonstrated with BA 44 activated for center-embedding and for sentences involving movement, and BA 45 selectively adapted to movement (Santi & Grodzinsky 2010). This finding is in line with the results reported by Makuuchi *et al.* (2009) for embedding and by Santi & Grodzinsky (2007) for movement.<sup>1</sup>

In the study by Makuuchi *et al.* (2009), working memory was neurally segregated from processing center-embedding. The latter recruited BA 44, whereas working memory necessary to bind the respective A and B elements during processing recruited the inferior frontal sulcus located dorsally to Broca's area. This is in line with studies that report phonological processes and phonologically-based working memory processes to activate "the dorsal aspect of the inferior frontal gyrus near the inferior frontal sulcus" (Poldrack *et al.* 1999; see also Vigneau *et al.* 2006).

Thus, the activation data reported here point towards a functional subdivision in the inferior frontal cortex with respect to different computational sub-components necessary to deal with syntactically complex recursive structures.

## 5. Processing Complex Hierarchy in a Non-Language Domain I: Visual-Spatial Event Sequences

When considering Broca's area as a brain region supporting the processing of complex structural hierarchies, the question arises whether this function is domain-specific or not. A direct way to approach this question is to investigate the processing of a hierarchical structure which matches that of the artificial grammars on syllable processing in a non-language domain.

We therefore conducted an fMRI study on the processing of hierarchical structures in a non-language domain (Bahlmann *et al.* 2009) using sequence structures just like those in the prior language studies. Category A and B elements were abstract visual stimuli whose membership was indicated by shape and texture. The dependency between A and B elements was encoded by rotation of the respective nonsense shape (see Figure 5).

---

<sup>1</sup> Note, that the statement that Broca's area supports the processing of complex hierarchical structures does not speak against the claim that Broca's area may also subserve the processing of non-hierarchical sequences (Pettersson *et al.* 2010). Except for the first study reviewed here all findings stem from a direct comparison between a complex hierarchical condition with a condition which involves a dependency between adjacent elements. Thus Broca's area is shown to increase its activity as a function of increasing hierarchical complexity.



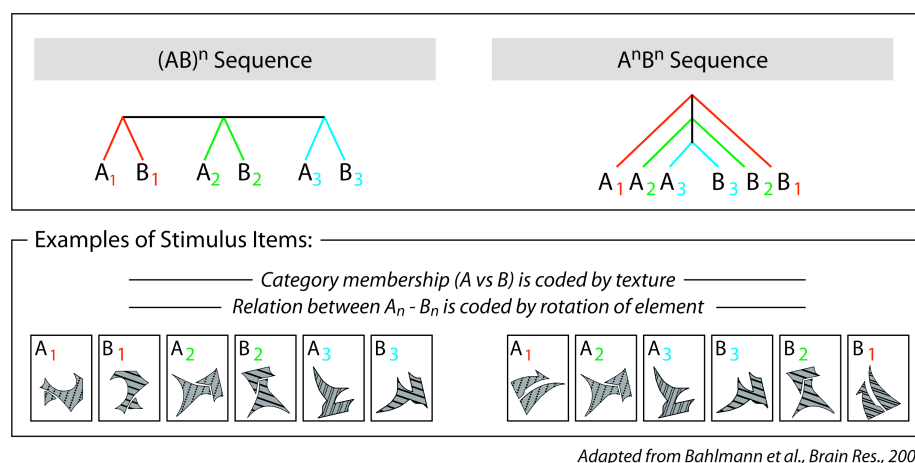


Figure 5: Processing of hierarchy in visuo-spatial event sequences. Top: Schematic view of the two structures. Bottom: Examples of stimuli. The relation between dependent elements is defined by rotation ( $B$  item has the identical shape as  $A$  item, but is spatially rotated). Dependency is color-coded (red, green, blue).

Processing of visual event-sequences in general (adjacent and hierarchical dependencies) activated the bilateral parietal lobe. A main effect of hierarchy was found for a whole brain analysis in the left pre-central gyrus (BA 6), the right pre-supplementary motor area and the right caudate. A hypothesis-driven region of interest analysis in BA 44 defined by a cytoarchitectonic probability map of area 44 (Amunts *et al.* 1999), however, revealed an increase of activation in BA 44 as a function of structural hierarchy (see Table 1 below and Figure 4 above). These data suggest that parts of the parietal cortex and pre-SMA together with BA 6 and BA 44 constitute the processing network for structured visual event sequences, and that BA 44/6 are involved when processing hierarchical dependencies.

From the present experiment in conjunction with those reported above, we may conclude that Broca's area receives its domain-specificity as a part of a particular neural network which differs from domain to domain. For example, Broca's area in a network together with the posterior superior temporal cortex subserves the processing of hierarchically complex natural language sentences, whereas Broca's area as part of a larger network involving the pre-motor cortex, the pre-SMA and parietal regions subserves the processing of non-linguistic visual-spatial event sequences.

The natural language experiment by Makuuchi *et al.* (2009) most directly indicates the BA 44 is part of the neural basis of linguistic recursion. The left posterior superior and middle temporal cortex seem to come into play when processing natural language sentences which require the assignment of thematic and semantic relations (Bornkessel *et al.* 2005, Snijders *et al.* 2009, Newman *et al.* 2010).

The present results for the non-language domain indicate that the view that Broca's area supports the processing of syntactic hierarchy in language does not preclude the involvement of Broca's area in other processing domains, be it the processing of visual-event sequences (Bahlmann *et al.* 2009), the processing of action sequences (e.g., Pulvermüller & Fadiga 2010), the processing of abstract action rules (e.g., Badre *et al.* 2010), or the processing of hierarchically ordered

control signals (e.g., Koechlin & Summerfield 2007). In these cases, however, Broca's area is part of a different neural network than the one observed for language processing. The view that Broca's area receives its specificity for syntactic processes as part of a specific network has previously been discussed in the literature (Friederici 2002, Marcus *et al.* 2003, Friederici 2006, Petersson *et al.* 2010).

## 6. Processing Complex Hierarchy in a Non-Language Domain II: Mathematical Formulae

Before a general conclusion with respect to the relation between Broca's area and the processing of complex structural hierarchy can be drawn, consideration needs to be given to whether the assumed relation also holds for hierarchies that do not mimic as the embedded structure used in the previous study. It has been proposed that recursion as assumed for language might also underlie mathematics and the processing of mathematical formulae (Hauser *et al.* 2002, Fitch 2010).

The goal of the next experiment was to see whether Broca's area is involved in the processing of structural hierarchy in mathematical formulae (Friedrich & Friederici 2009). There is no doubt that in mathematics, a person familiar with the respective rules can make grammaticality judgements such as evaluating the correctness of a recursive structure. This experiment was, therefore, conducted with mathematicians. The formulae used in this experiment had either a hierarchical structure or a "linear" structure (see Figure 6). The hierarchical structure of these formulae was not primarily determined by embeddedness, but by the number of levels in the tree structure.

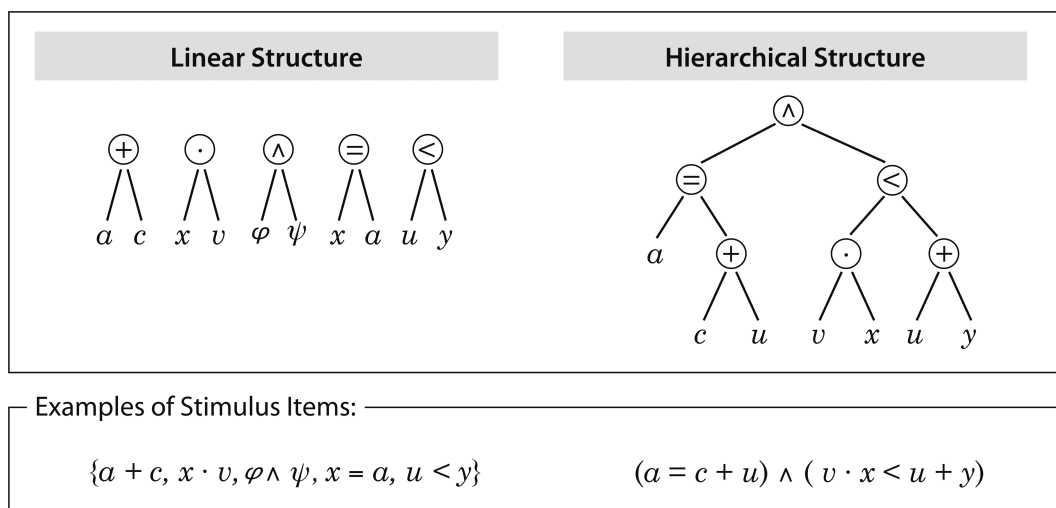


Figure 6: Processing of mathematical formulae. Top: Schematic view of the two structures. Nodes (circled) indicate the operator. Bottom: Examples of stimulus items.

It should be noted that hierarchy in mathematical formulae tends to differ from hierarchy in natural languages. Language structures are usually asymmetric whereas mathematical structures need not necessarily be so, as exemplified in Figures 3b and 6. While Figure 3b displays the linguistic description of a center-embedded sentence used in the natural language study (Makuuchi *et al.* 2009), Figure 6 shows the structure of mathematical formulae used in the mathematical study (Friedrich & Friederici 2009). Crucially, the nodes in the mathematical formulae (circled in Figure 6) contain an operator indicating the operation between the respective elements, i.e. = means ‘equals’, < means ‘larger than’, etc. These operators require that the two elements under the respective node must be put into a logical relation. This may require the activation of additional or even different brain regions than those observed in the processing of the hierarchical structures in the previous experiments.

The formulae used as stimuli in the mathematical study did not contain numbers, in order to abstract from the issue of numerosity and related number-based calculation processes. The formulae presented in the fMRI experiment were either correct or incorrect. Participants were students of mathematics and physics and were therefore highly familiar with mathematical formula processing. They were required to make judgements regarding the correctness of the visually presented formulae. Whole head analysis of the brain imaging data for the processing of these mathematical formulae revealed a clear effect of hierarchy in left BA 47 bordering BA 45 and in parietal regions, as well as the right precuneus (see Table 1 below and Figure 7).

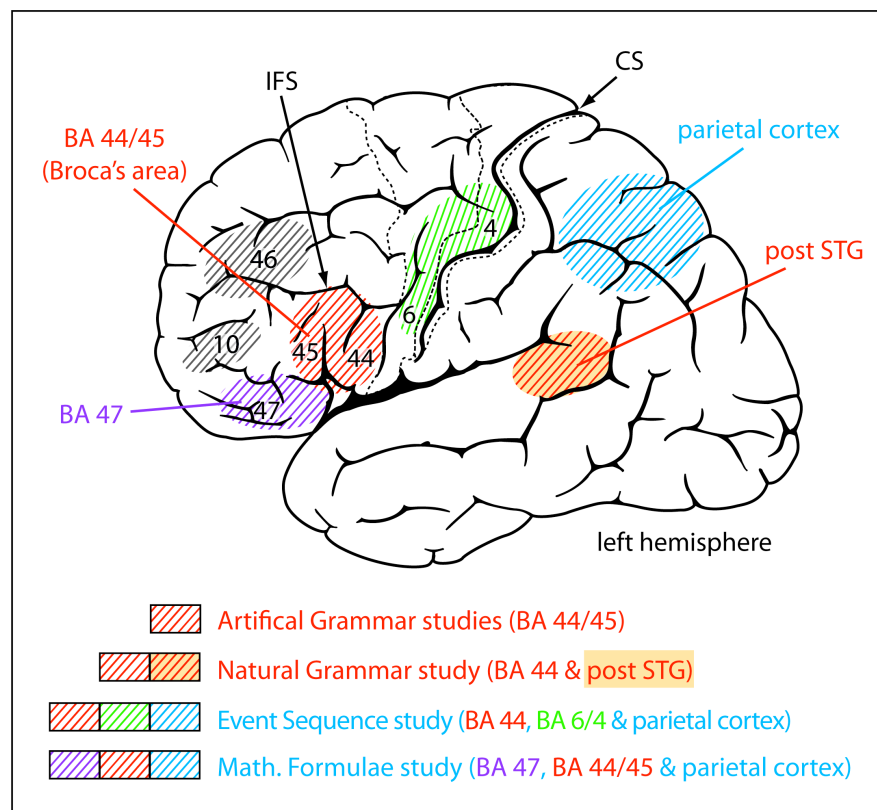


Figure 7: Schematic view of the activation pattern for the main effect of hierarchy in the language and non-language domains. For, explanation of activation for grammar studies, see Figure 4. For the visuo-spatial event sequence study, the main effect of hierarchy was found in the precentral gyrus (BA 6/4); a main effect of hierarchy in Broca's area (BA 44/45) was only found in a region of interest analysis (Bahlmann et al. 2009). For the mathematical formulae study, the main effect of hierarchy was found in BA 47 bordering BA 45 as well as in the medial frontal gyri (BA 10) and the most dorsal part of middle frontal gyrus (BA 6) (not depicted in the figure). In addition, a hierarchy effect was found in the parietal lobule bilaterally. For details, see Friedrich & Friederici (2009).

Key: BA = Brodmann Area; CS = central sulcus; IFS = inferior frontal sulcus; preSMA = pre-supplementary motor area; STG = superior temporal gyrus.

Given the previous analyses conducted by Bahlmann *et al.* (2009), which revealed an involvement of Broca's area for the processing of hierarchical structures in the visuo-spatial domain only in a region of interest analysis, we computed a similar analysis for the mathematical domain for the present article. This region of interest analysis for the voxels defined by the cytoarchitectonic probability map of area 44 by Amunts *et al.* (1999) revealed an effect of hierarchy for the correct formulae ( $p < .05$ ) (see Table 1 below). Thus, BA 44 partly supports the processing of hierarchy in mathematical formulae, although the crucial area which most strongly subserves this process in the prefrontal region is located more anteriorly, namely BA 47 bordering on BA 45.

The obvious difference between hierarchical structures used in the mathematical formulae processing study (Friedrich & Friederici 2009) and the embedded structures used in the other studies (Bahlmann *et al.* 2008, 2009, Makuuchi *et al.* 2009) is that in the former, the nodes in the syntactic tree are operators calling for logical processes. Thus one of the crucial aspects in the comparison of hierarchically structured and linear mathematical formulae may be that for a successful judgment of the logical relations indicated by the operators, increased logical-semantic processes are necessary, recruiting BA 47 bordering on BA 45. This interpretation is in line with the view that BA 47 (and the anterior part of 45) mainly supports semantic processes, whereas the more posterior region, namely BA 44 (and the posterior part of BA 45) mainly subserves syntactic processes during language processing (see Bookheimer 2002, Friederici 2002, Hagoort 2005, Vigneau *et al.* 2006)<sup>2</sup>.

In the context of cognitive control models of the prefrontal cortex (PFC), which assume a posterior-to-anterior gradient with a recruitment of more anterior portions of the PFC as hierarchies become more complex (for a recent review, see Botvinick 2008), the present data could make an interesting contribution.

## 7. Hierarchy in the Prefrontal Cortex

In order to see how far the present set of studies can be interpreted in the context of a general model of the PFC for the processing of hierarchies we compare the

<sup>2</sup> Note that a novel receptorarchitectonic study suggests a neuroanatomical subdivision of BA 45 into an anterior (area 45a) and a posterior (area 45p) part (Amunts *et al.* 2010). It seems likely that the receptorarchitectonic division of BA 45 is also functionally relevant.

different studies and the receptive activation in the PFC directly. Please note that the first Artificial Grammar Study I (Friederici *et al.* 2006a) is not included, as a direct test for the hierarchy effect was not possible due to the fact that the two grammar types (complex vs. simple) was a between-group factor. The other studies, with their location of the main effect of hierarchy, are listed in Table 1. The second artificial grammar study (Bahlmann *et al.* 2008) and the natural language study (Friederici *et al.* 2009, Makuuchi *et al.* 2009) revealed a main effect of hierarchy in BA 44. For the two non-language studies, a main effect of hierarchy in BA 44 was only seen in a ROI analysis. In the whole brain analysis for the visuo-spatial event sequences, a main effect of hierarchy was observed in the left precentral gyrus, the right pre-SMA and the right caudate, and for mathematical formulae in BA 47 and 45a.

Study	BA	X	Y	Z
Artificial Grammar II Bahlmann <i>et al.</i> (2009) WH	44	-46	5	16
Natural Grammar Friederici <i>et al.</i> (2009) WH	44	-45	6	21
Makuuchi <i>et al.</i> (2009) ROI BA 44	44	n.a.		
Visuo-spatial sequence Bahlmann <i>et al.</i> (2009) WH	6/4	-50	-8	33
ROI BA 44	44	n.a.		
Mathematical Formulae Friedrich & Friederici (2009) WH	45 47 10	-47 -38 -38	19 52 52	6 -3 -3
ROI BA 44 (conducted for the present article)	44	n.a.		

Table 1: Anatomical areas, Brodmann Areas (BA) mean Talairach coordinates (X, Y, Z) for significant effect of hierarchy in left prefrontal cortex, WH = whole head analysis, ROI = region of interest analysis based on cytoarchitectonic definition of BA 44 with a probability of 30% (Amunts *et al.* 1999), for which Talairach coordinates are not applicable (n.a.).

Current models of the prefrontal cortex (PFC) assume a posterior-to-anterior gradient as the neural basis of hierarchically organized behavior. The posterior-to-anterior dimension in the lateral PFC has been considered a key in the temporal integration of behavior (Fuster 1990). Alternative models proposed a posterior-to-anterior functional gradient for executive control in action selection (Koechlin *et al.* 2003, Koechlin & Summerfield 2007, Badre 2008, Badre *et al.* 2010). The posterior-to-anterior gradient goes from the premotor cortex (BA 6) located

in the posterior PFC, over the posterior dorsal lateral PFC (BA 44/45) to the anterior dorsolateral PFC (BA 46/47) and further to the polar portion of the PFC (BA 10), with more abstract, hierarchically structured processes recruiting more anterior regions (Koechlin & Jubault 2006, Badre 2008). It should be noted that both these latter theories lay no direct claim as to whether the models hold for the processing of hierarchical sequences in the language domain (but see Koechlin & Jubault 2006). If they did, these theories would be compatible with the studies discussed here only under a view assuming that the processing of mathematical formulae requires more executive control than the processing of linguistic structures. If, however, the crucial parameter according to which the prefrontal cortex is functionally organized is 'complexity of hierarchy' of a given stimulus, the present data are not fully compatible with such theories, since the 'complexity of hierarchy' of the stimulus does not fully determine the localization of the activation in the prefrontal cortex.

It seems that the posterior-to-anterior gradient correlates with qualitatively different computations required. The computation of mathematical formulae, which include logical operations indicated by operators at the structural nodes, relies on the more anterior ventral part of the IFG, namely BA 47/45a, whereas the computation of hierarchical structures in natural language is localized in more posterior regions of the IFG, namely in BA 44/45p. Complexity of hierarchy of a given sequence does not fully determine the localization in the prefrontal cortex, as the structures tested in the natural language experiment are quite complex (for a linguistic description of such a sentence see Figure 3b). These linguistic structures, however, only recruit areas located in the most posterior part of the IFG, i.e. BA 44, which, according to the models above, are responsible for the processing of less complex hierarchies. Note, that other studies in the literature often report syntax-related activation in BA 45 (Ben-Shachar *et al.* 2004, Bornkessel *et al.* 2005, Santi & Grodzinsky 2007, 2010, Snijders *et al.* 2009, Pallier *et al.* 2011). It remains to be determined whether the cytoarchitecturally different regions BA 44 and BA 45 can be functionally separated or whether the receptor-architectonic separation between the more anterior portion of the IFG covering area 47/45a and the more posterior portion covering area 44/45p and is functionally relevant. Independent of this fine grained neuroanatomical distinction the present data show that highly hierarchically complex language structures can be dealt with by the posterior IFG, whereas the processing of hierarchical mathematical formulae requiring logical reasoning recruits more anterior brain regions.<sup>3</sup>

One important aspect of the processing of mathematical formulae as compared to language processing may be that even for mathematicians, the processing of mathematical formulae could be less automatic, requiring more cognitive control than the processing of language hierarchies. The data available do not allow us to ultimately decide to what extent the observed differences in the PFC activation are driven by the difference in the processing domains, as it is conceivable that familiarity with language structures is considerably greater than with mathematical formulae even in mathematicians.

---

<sup>3</sup> For a discussion of the function of Broca's in language and its role in Broca's aphasia, see Grodzinsky & Amunts (2006) and the contributions therein.

If valid, this interpretation would call upon a view suggesting two parallel systems dealing with hierarchical structures, one which following the posterior-to-anterior gradient is determined by the degree of cognitive control leading to activation in the anterior PFC (BA 47/45a and 10) for highly complex sequences in different domains, and one which is confined to the posterior IFG (BA 44/45p) and which in the adult brain efficiently deals with highly complex hierarchically structured language sequences. When language processes are less automatic as during first and second language acquisition more anterior regions of the PFC have to be recruited even for processing local phrase structure hierarchies (Rüschmeyer *et al.* 2005, Brauer & Friederici 2007).

## 8. Conclusions

Language processing in adults is highly automatic and does not appear to be very challenging for the brain, even when the sequences to be processed are hierarchically complex. One intriguing conclusion is that humans are predetermined to compute linguistic recursion, with BA 44/45p being the neural correlate of this showing its functional primacy in the adult brain after long language exposure. Based on the studies discussed here, we propose that there are two different computational systems in the lateral PFC dealing with hierarchical structures: one system determined by cognitive control that follows the posterior-to-anterior gradient and one system confined to Broca's area which is able to process complex hierarchies in language efficiently.

## References

- Amunts, Katrin, Marianne Lenzen, Angela D. Friederici, Axel Schleicher, Patricia Morosan, Nicola Palomero-Gallagher & Karl Zilles. 2010. Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biology* 8, e100489.
- Amunts, Katrin, Axel Schleicher, Uli Bürgel, Hartmut Mohlberg, Harry B.M. Uylings & Karl Zilles. 1999. Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology* 412, 319–341.
- Badre, David. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences* 12, 193–200.
- Badre, David, Andrew S. Kayser & Mark D'Esposito. 2010. Frontal cortex and the discovery of abstract action rules. *Neuron* 66, 315–326.
- Bahlmann, Jörg, Ricarda I. Schubotz & Angela D. Friederici. 2008. Hierarchical sequencing engages Broca's area. *NeuroImage* 42, 525–534.
- Bahlmann, Jörg, Ricarda I. Schubotz, Jutta L. Mueller, Dirk Koester & Angela D. Friederici. 2009. Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research* 1298, 161–170.
- Ben-Shachar, Michal, Dafna Palti & Yosef Grodzinsky. 2004. Neural correlates of

- syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage* 21, 1320–1336.
- Bornkessel, Ina, Stefan Zysset, Angela D. Friederici, Yves D. von Cramon & Matthias Schlesewsky. 2005. Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage* 26, 221–223.
- Botvinick, Matthew M. 2008. Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences* 12, 201–208.
- Brauer, Jens & Angela D. Friederici. 2007. Functional neural networks of semantic and syntactic processes in the developing brain. *Journal of Cognitive Neuroscience* 19, 1609–1623.
- Bookheimer, Susan. 2002. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience* 25, 151–188.
- Cooke, Ayanna, Edgar B. Zurif, Christian DeVita, David Alsop, Phyllis Koenig, John Detre, James Gee, Maria Pinango, Jennifer Balogh & Murray Grossman. 2001. Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping* 15, 80–94.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Fitch, Tecumseh W. 2010. Three meanings of recursion: Key distinctions for biolinguistics. In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*, 73–90. Cambridge: Cambridge University Press.
- Friederici, Angela D. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78–84.
- Friederici, Angela D. 2004. The neural basis of syntactic processes. In Michael S. Gazzaniga (ed.), *The Cognitive Neurosciences III*, 789–801. Cambridge, MA: MIT Press.
- Friederici, Angela D. 2006. Broca's area and the ventral premotor cortex in language: Functional differentiation and specificity. *Cortex* 42, 472–475.
- Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schubotz & Alfred Anwander. 2006a. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America* 103, 2458–2463.
- Friederici, Angela D., Christian J. Fiebach, Matthias Schlesewsky, Ina D. Bornkessel & Yves D. von Cramon. 2006b. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex* 16, 1709–1717.
- Friederici, Angela D., Michiru Makuuchi & Jörg Bahlmann. 2009. The role of the posterior temporal cortex in sentence comprehension. *NeuroReport* 20, 563–568.
- Friedrich, Roland & Angela D. Friederici. 2009. Mathematical logic in the human brain. *PLoS ONE* 4, e5599.
- Fuster, Joaquín M. 1990. Behavioral electrophysiology of the prefrontal cortex of the primate. *Progress in Brain Research* 85, 313–324.
- Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.



- Grodzinsky, Yosef & Katrin Amunts (eds.). 2006. *Broca's Region*. Oxford: Oxford University Press.
- Grodzinsky, Yosef & Angela D. Friederici. 2006. Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology* 16, 240–246.
- Hagoort, Peter. 2005. On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences* 9, 416–423.
- 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.
- Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language. *Cognition* 97, 211–225.
- Jonides, John, Edward E. Smith, Christy Marshuetz, Robert A. Koeppe & Patricia A. Reuter-Lorenz. 1998. Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Science of the United States of America* 95, 8410–8413.
- Koechlin, Etienne, Chryssèle Ody & Frédérique Kouneiher. 2003. The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185.
- Koechlin, Etienne & Christopher Summerfield. 2007. An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences* 11, 229–235.
- Koechlin, Etienne & Thomas Jubault. 2006. Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Makuuchi, Michiru, Jörg Bahlmann, Alfred Anwander & Angela D. Friederici. 2009. Segregating the core computation of human language from working memory. *Proceedings of the National Academy of Science of the United States of America* 106, 8362–8367.
- Markus, Gary F., Athena Vouloumanos & Ivan A. Sag. 2003. Does Broca's play by the rules? *Nature Neuroscience* 6, 651–652.
- Newman, Sharlene D., Toshikazu Ikuta & Thomas Burns Jr. 2010. The effect of semantic relatedness on syntactic analysis: An fMRI study. *Brain and Language* 113, 51–58.
- Pallier, Christophe, Anne-Dominique Devauchelle & Stanislas Dehaene. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America* 108, 2522–2527.
- Perruchet, Pierre & Arnaud Rey. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin and Review* 12, 307–313.
- 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]
- Poldrack, Russel A., Anthony D. Wagner, Matthew W. Prull, John E. Desmond, Gary H. Glover & John D. E. Gabrieli. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10, 15–35.
- Pulvermüller, Friedemann & Luciano Fadiga. 2010. Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*

- 11, 351–360.
- Rüschemeyer, Shirley-Ann, Christian J. Fiebach, Vera Kempe & Angela D. Friederici. 2005. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping* 25, 266–286.
- Sanides, Friedrich. 1962. Entwicklungsprinzipien des menschlichen Stirnhirns. *Naturwissenschaften* 49, 160–161.
- Santi, Andrea & Yosef Grodzinsky. 2007. Working memory and syntax interact in Broca's area. *NeuroImage* 37, 8–17.
- Santi, Andrea & Yosef Grodzinsky. 2010. fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage* 51, 1285–1293.
- Smith, Edward E. & John Jonides. 1998. Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Science of the United States of America* 95, 12061–12068.
- Smith, Edward E. & John Jonides. 1999. Storage and executive processes in the frontal lobes. *Science* 283, 1657–1661.
- Snijders, Tineke M., Theo Vosse, Gerard Kempen, Jos J. A. van Berkum, Karl Magnus Petersson & Peter Hagoort. 2009. Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex* 19, 1493–1503.
- Vigneau, Mathieu, Virginie Beaus Cousin, Pierre-Yves Herve, Hugues Duffau, Fabrice Crivello, Olivier Houde, Bernad Mazoyer & Nathalie Tzourio-Mazoyer. 2006. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* 30, 1414–1432.
- 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.

Angela D. Friederici  
 Max Planck Institute for Human Cognitive  
 and Brain Sciences  
 Stephanstr. 1a  
 04103 Leipzig  
 Germany  
[angelafr@cbs.mpg.de](mailto:angelafr@cbs.mpg.de)

& Stanford University Center for Advanced  
 Study in the Behavioral Science, USA

Roland Friedrich  
 Humboldt University  
 Berlin Mathematical School  
 Strasse des 17. Juni 136  
 10623 Berlin  
 Germany  
[rolandf@mathematik.hu-berlin.de](mailto:rolandf@mathematik.hu-berlin.de)

Jörg Bahlmann  
 University of California  
 Helen Wills Neuroscience Institute  
 3210F Tolman Hall MC 3192  
 Berkeley, CA 94720  
 USA  
[bahlmann@berkeley.edu](mailto:bahlmann@berkeley.edu)

Michiru Makuuchi  
 Max Planck Institute for Human Cognitive  
 and Brain Sciences  
 Stephanstr. 1a  
 04103 Leipzig  
 Germany  
[makuuchi@cbs.mpg.de](mailto:makuuchi@cbs.mpg.de)

# Implicit Artificial Syntax Processing: Genes, Preference, and Bounded Recursion

Vasiliki Folia, Christian Forkstam, Martin Ingvar,  
Peter Hagoort & Karl Magnus Petersson

The first objective of this study was to compare the brain network engaged by preference classification and the standard grammaticality classification after implicit artificial syntax acquisition by re-analyzing previously reported event-related fMRI data. The results show that preference and grammaticality classification engage virtually identical brain networks, including Broca's region, consistent with previous behavioral findings. Moreover, the results showed that the effects related to artificial syntax in Broca's region were essentially the same when masked with variability related to natural syntax processing in the same participants. The second objective was to explore CNTNAP2-related effects in implicit artificial syntax learning by analyzing behavioral and event-related fMRI data from a subsample. The CNTNAP2 gene has been linked to specific language impairment and is controlled by the FOXP2 transcription factor. CNTNAP2 is expressed in language related brain networks in the developing human brain and the FOXP2-CNTNAP2 pathway provides a mechanistic link between clinically distinct syndromes involving disrupted language. Finally, we discuss the implication of taking natural language to be a neurobiological system in terms of bounded recursion and suggest that the left inferior frontal region is a generic on-line sequence processor that unifies information from various sources in an incremental and recursive manner.

*Keywords:* artificial grammar learning; artificial language; Broca's region; CNTNAP2; fMRI; FOXP2; genes, grammaticality classification; natural language; preference classification; syntax

## 1. Introduction

Human languages are characterized by universal "design features" (Hockett 1963, 1987): discreteness, arbitrariness, productivity, and the duality of patterning

---

This work was supported by the Max Planck Institute for Psycholinguistics, the Donders Institute for Brain, Cognition and Behaviour, the Fundação para a Ciência e Tecnologia (PTDC/PSI-PCO/110734/2009; IBB/CBME, LA, FEDER/POCI 2010), the Stockholm Brain Institute, Vetenskapsrådet, the Swedish Dyslexia Foundation, the Hedlunds Stiftelse, and the Stockholm County Council (ALF, FoUU).



(i.e. elements at one level are combined to construct elements at another). Somehow these properties arise from the way the human brain processes, develops, and learns, in interaction with its environment. The human capacity for language and communication is subserved by a network of brain regions that collectively instantiate the phonological, syntactic, semantic, and pragmatic operations necessary for adequate language production and comprehension. During normal language processing, phonology, syntax, and semantics operate in close temporal and spatial contiguity in the human brain. Therefore the artificial grammar learning (AGL) paradigm has been used to create a relatively uncontaminated window onto the neurobiology of syntax. Artificial syntax learning paradigms thus makes it possible to investigate structured sequence processing relatively independent of, for example, semantics and phonology (Petersson *et al.* 2004, 2010). In addition, artificial syntax learning has been used for cross-species comparisons in an attempt to establish the uniquely human component of the language faculty (Hauser *et al.* 2002, Fitch & Hauser 2004, O'Donnell *et al.* 2005, Gentner *et al.* 2006, Saffran *et al.* 2008).

Artificial syntax learning paradigms have been widely employed to study different aspects of natural language acquisition (Gómez & Gerken 2000, Folia *et al.* 2010), though it was originally implemented to investigate the underlying implicit sequence learning mechanism, which is presumably shared with natural language learning (Reber 1967) as well as other situations in which new skills are acquired (e.g. Misyak *et al.* 2009, 2010a, 2010b). The neurobiology of implicit sequence learning as assessed by artificial syntax acquisition have been investigated by means of functional neuroimaging (e.g. Petersson *et al.* 2004, 2010, Forkstam *et al.* 2006), brain stimulation (Uddén *et al.* 2008, 2011, de Vries *et al.* 2010), and agrammatic aphasics (Christiansen *et al.* 2010), and generally involve frontostriatal circuits (Packard & Knowlton 2002, Ullman 2004; note that implicit learning is sometimes referred to as procedural learning, and vice versa), which are also involved in the acquisition of natural syntax (Ullman 2004). More specifically, recent functional neuroimaging (e.g. Petersson *et al.* 2004, 2010, Forkstam *et al.* 2006) and brain stimulation research (Uddén *et al.* 2008, 2011, de Vries *et al.* 2010), have identified some of the brain regions involved, including repeatedly showing that Broca's region, a brain region involved in natural syntax processing, is also involved in artificial syntax processing. Indeed, the breakdown of syntax processing in agrammatic aphasia is associated with impairments in artificial syntax learning (Christiansen *et al.* 2010). Moreover, Conway & Pisoni (2008) found that individual variability in implicit sequence learning correlated with language processing. Supportive evidence also comes from a recent study by Misyak *et al.* (2010a), who found that individual differences in learning non-adjacent dependencies, assessed by non-linguistic implicit sequence learning, correlate with the processing of natural language sentences containing complex non-adjacent dependencies. This supports the hypothesis that artificial grammar learning paradigm taps into implicit structured sequence learning and artificial syntax processing, and thus provides a useful way to investigate aspects of natural language processing. Thus, there is a growing body of evidence that language acquisition and language processing, both a natural and artificial setting, is mediated by implicit sequence learning and structured sequence proces-

sing mechanisms, respectively.

The implicit artificial syntax learning paradigm allows for a systematic investigation of aspects of structural acquisition from grammatical examples without providing explicit feedback, teaching instruction, or engaging the subjects in explicit problem solving (Forkstam *et al.* 2006, 2008, Folia *et al.* 2008). These acquisition conditions resemble, in certain important respects, those found in natural-language development with respect to syntax acquisition (Chomsky & Miller 1963: 275–276). Generally, artificial grammar learning paradigms consist of acquisition and test phases. In the acquisition phase, participants are exposed to an acquisition sample generated from a formal grammar. In the standard version, subjects are informed that the sequences were generated according to a complex set of rules after acquisition, and are asked to classify novel sequences as grammatical or not, based on their immediate intuitive impression (i.e. guessing based on gut-feeling). A well-replicated and robust finding in this paradigm is that subjects perform well above chance after several days of implicit acquisition; they do so on regular (e.g. Stadler & Frensch 1998, Folia *et al.* 2008, Forkstam *et al.* 2008) as well as non-regular grammars, including those that generate context-free and context-sensitive non-adjacent dependencies (Uddén *et al.* 2009).

In this study, we investigate an implicit preference AGL paradigm with several days of acquisition. During the implicit acquisition period, participants were exposed to grammatical sequences only in a cover task based on the structural mere-exposure effect (Zajonc 1968, Zizak & Reber 2004, Folia *et al.* 2008, Forkstam *et al.* 2008). The structural mere-exposure effect refers to the finding that repeated exposure to a stimulus created by a certain rule system, induces an increased preference for novel stimuli conforming to the same underlying system (Zizak & Reber 2004). To this end, we exposed the participants to a simple right-linear unification grammar — a grammar that generates right-linear phrase structures (Vosse & Kempen 2000, Hagoort 2005, Petersson *et al.* 2010). During the acquisition period, spanning five days, subjects were exposed to syntactically well-formed consonant sequences and no performance feedback was provided. On the last day a preference classification test was administered in which new sequences were presented. Previously, the implicit preference AGL paradigm has been characterized exclusively in behavioural terms (e.g. Manza & Bornstein 1995, Zizak & Reber 2004, Folia *et al.* 2008, Forkstam *et al.* 2008). Here we first review the outcome of implicit artificial syntax acquisition from an event-related fMRI study (Folia *et al.* 2011). Then we compare the brain network engaged by preference classification and the standard grammaticality classification after implicit artificial syntax acquisition from a previously reported event-related fMRI results on the standard grammaticality classification paradigm in the same subjects (Petersson *et al.* 2010). In addition, we investigate the common overlap between artificial and natural syntax processing by masking the non-grammatical (NG) vs. grammatical (G) effect observed in preference classification with the natural-syntax-related variability in the same subjects (Folia *et al.* 2009). Consistent with the hypothesis of implicit utilization of acquired structural knowledge as well as previous behavioral results (Forkstam *et al.* 2008), which showed that subjects perform qualitatively identical on preference and grammaticality classification, we found that the brain network subserving preference

classification during artificial syntax processing engaged Broca's region centered on Brodmann's areas (BA) 44 and 45, and did not differ from those observed during grammaticity classification. This strengthens the notion that preference and grammaticity classification in the implicit artificial syntax learning are essentially equivalent (Forkstam *et al.* 2008). Finally, based on these event-related fMRI data (Petersson *et al.* 2010, Folia *et al.* 2011), we took advantage of the fact that a subsample of our participants was part of the Brain Imaging Genetics (BIG) project at the Donders Centre for Cognitive Neuroimaging and the Department of Human Genetics of the Radboud University Nijmegen. This allowed us to explore the potential role of the CNTNAP2 gene in artificial syntax acquisition/processing at the behavioral as well as the brain level.

Relatively recently, language research has started to investigate the role of genes in language (Enard *et al.* 2002, Vargha-Khadem *et al.* 2005, Bishop 2009, Konopka *et al.* 2009). For example, mutations in the FOXP2 gene result in a complex symptomatology, called developmental verbal dyspraxia, which includes difficulties with learning and producing sequences of oral movements relevant for speech, as well as impairments in morphosyntactic aspects of language processing (Lai *et al.* 2001, Watkins *et al.* 2002, MacDermot *et al.* 2005). FOXP2 is a gene that codes for the transcription factor (a protein) *foxp2* which regulates gene expression during development. This means that *foxp2* controls the production of other proteins coded for by other genes. Transcription factors and their genes make up complex gene regulatory networks, which control many complex biological processes, including ontogenetic development (Davidson *et al.* 2002, Davidson 2006, Alberts *et al.* 2007). Moreover, functional neuroimaging studies of the KE family (with a protein-truncating FOXP2 mutation; Lai *et al.* 2001), have demonstrated structural and functional abnormalities in brain regions related to language (Vargha-Khadem *et al.* 2005). The CNTNAP2 gene has been linked to specific language impairment (SLI) and the FOXP2–CNTNAP2 pathway provides a mechanistic link between clinically distinct syndromes involving disrupted language (Vernes *et al.* 2008). The CNTNAP2 gene is controlled (down-regulated) by the *foxp2* transcription factor (Vernes *et al.* 2008). CNTNAP2 codes for a neural trans-membrane protein, which belongs to neurexin superfamily (Poliak *et al.* 1999) and it has been shown that, in the developing human brain, the expression of CNTNAP2 is relatively increased in fronto-temporal-subcortical brain networks (Alarcón *et al.* 2008). In particular, the CNTNAP2 expression is enriched in frontal brain regions in humans, but not in mice or rats (Abrahams *et al.* 2007). A recent study investigated the effects of a common single nucleotide polymorphism (SNP) RS7794745 in the CNTNAP2 gene on the brain response during language comprehension (Snijders *et al.* 2011). This study found both structural and functional brain differences in language comprehension related to the same SNP sub-grouping used in this study.

Finally, we note that an artificial grammar represents a formal specification of the mechanism that generates, for example, specific structural or sequence regularities (e.g., various types of local or non-adjacent dependencies). From this point of view, an artificial syntax is a formal language (Davis *et al.* 1994) and artificial syntax learning is an experimental model to investigate various (any) generative mechanism independent of other aspects of a language (cf. the

introduction of Petersson *et al.* 2004). As noted above, artificial syntax learning can be used as an experimental tool to investigate the processing properties of Broca's region, a central node in the brain network for natural syntax processing. In this context, we take the view that natural and artificial syntax processing share a common abstraction — structured sequence processing. Clearly, any particular artificial grammar cannot instantiate all phenomena found in natural syntax. Rather, in experimental work it is necessary to focus on some particular aspect of syntax, which is also the case for experimental work on natural language syntax. Artificial syntax learning thus provides a window onto the neurobiology of syntax, in the sense that artificial syntax learning allows us to investigate the computational properties of Broca's region. In the Discussion section, we return to some issues related to the present the Chomsky hierarchy and recursive processing from the point of view that natural language is a neurobiological system.

## 2. Materials and Methods

### 2.1. Participants

Here we briefly describe the relevant background of the material and methods used by Vasiliki Folia and colleagues (Folia *et al.* 2008, 2011, Petersson *et al.* 2010) as they apply to this study. Thirty-two healthy right-handed Dutch university students were recruited in the study (16 females, mean age  $\pm$  SD = 22  $\pm$  3 years; mean years of education  $\pm$  SD = 16  $\pm$  2). None of the subjects used any medication, had a history of drug abuse, head trauma, neurological or psychiatric illness, or a family history of neurological or psychiatric illness. All subjects had normal or corrected-to-normal vision. Written informed consent was obtained from all participants according to the Declaration of Helsinki as well as from the local medical ethics committee. Of the thirty-two participants, twelve were already included in the BIG database at the Donders Centre for Cognitive Neuroimaging and the Department of Human Genetics of the Radboud University Nijmegen (5 females, mean age  $\pm$  SD = 22  $\pm$  2 years; mean years of education  $\pm$  SD = 16  $\pm$  2) and typed for the single nucleotide poly-morphism (SNP) RS7794745 (with a breakdown on AA:AT:TT of 4:6:2). Because of the few TT-carriers, we pooled all T-carriers into one group of TT- and AT-carriers and analyzed the data in the T (N = 8) and nonT (N = 4) groups.

### 2.2. Stimulus Material

We used a simple right-linear unification grammar (Petersson *et al.* 2010) with the following vocabulary of terminal symbols (M, S, V, R, X) and lexicon of primitive trees (*treelets*)  $\{[s_1, [M, s_2]], [s_2, [S, s_2]], [s_2, [V, s_4]], [s_3, [X, s_2]], [s_3, [X, s_5]], [s_4, [R, s_3]], [s_4, [S, s_6]], [s_4, \#], [s_5, [R, s_5]], [s_5, [M, s_6]], [s_5, \#], [s_6, \#]\}$ . For a given lexical item (e.g.,  $[s_j, [T, s_k]]$ ),  $s_j$ ,  $s_k$  can be interpreted as syntactic control features and T as a surface feature. Within the unification framework (Vosse & Kempen 2000, Hagoort

2005, Petersson *et al.* 2010), an incoming sequence of surface symbols (e.g., MSV) initiates the retrieval of lexical items from the mental lexicon. As a result, they enter a unification space for on-line processing:  $[s_1, [M, s_2]], [s_2, [S, s_2]], [s_2, [V, s_4]] \dots$ , where two lexical items (e.g.,  $[s_i, [R, s_j]], [s_k, [Q, s_i]]$ ) unify (i.e. combine or merge) through a *unification operation*  $U$  if and only if  $s_j = s_k$  or  $s_i = s_i$ . This process is *incremental and recursive*. For example, if the structure  $U([s_1, [M, s_2]], [s_2, [S, s_2]]) = [s_1, [M, [s_2, [S, s_2]]]]$  is already present in the unification space when the lexical item  $[s_2, [V, s_4]]$  is retrieved, a larger combinatorial structure can be formed by the unification operation  $U([s_1, [M, [s_2, [S, s_2]]]], [s_2, [V, s_4]]) = [s_1, [M, [s_2, [S, [s_2, [V, s_4]]]]]]$ , and so on. The Unification operator works in the same way in all unification grammars. However, the structures generated by the Unification operator depend on the structure of the lexical items in any given grammar. In the present case, our grammar yields right-linear structures.

Folia *et al.* (2011) used a  $2 \times 2 \times 2$  factorial design including the factors instruction type (preference/grammaticality instruction), grammaticality status (grammatically correct/incorrect), and local subsequence familiarity (high/low ACS). The local subsequence familiarity (cf. Knowlton & Squire 1996, Meulemans & van der Linden 1997, Forkstam *et al.* 2006 for technical descriptions) is an associative measure of the superficial resemblance between classification sequences and the sequences in the acquisition set. The classification sequences with high ACS contain subsequences (bigrams and trigrams) that appear frequently in the acquisition set, while sequences with low ACS contain subsequences with a low frequency in the acquisition set. In total, 569 G sequences from the grammar, with a sequence length ranging from 5 to 12, were generated. For each item the frequency distribution of 2 and 3 letter chunks for both terminal and complete sequence positions was calculated. In this way, the associative chunk strength (ACS) was calculated for each item (cf. Knowlton & Squire 1996, Meulemans & van der Linden 1997, Forkstam *et al.* 2006). Next, for the acquisition set, 100 sequences representative, in terms of letter chunks, for the complete sequence set were randomly selected in an iterative way. In the next step, the NG sequences were created, derived from non-selected G sequences, by switching letters in two non-terminal positions. The NG sequences matched the G sequences in terms of both terminal and complete-sequence ACS (Forkstam *et al.* 2006, 2008). Finally, in an iterative procedure, we randomly selected two sets of 56 sequences each from the remaining G sequences, to serve as classification sets. The classification sets thus consisted of 25% grammatical/high ACS (HG); 25% grammatical/low ACS (LG); 25% non-grammatical/high ACS (HNG); and 25% non-grammatical/low ACS (LNG) sequences. See Appendix A below for example stimuli.

### 2.3. Experimental Procedures

During the acquisition sessions, subjects were presented with the 100 acquisition sequences (presentation order randomized for each acquisition session) and the task was an immediate short-term memory task serving as a cover task. Each sequence was centrally presented letter-by-letter on a computer screen (3–7 s corresponding to 5–12 terminal symbols; 300 ms presentation, 300 ms inter-symbol-interval) using the Presentation software (<http://nbs.neuro-bs.com>).



When the last letter in a sequence disappeared, subjects were instructed to reconstruct the sequence from memory and type it on a keyboard. No performance feedback was given, and only grammatical sequences were presented. The acquisition phase lasted approximately 20–40 minutes and took place over five consecutive days.

After the acquisition session on the last (5th) day of the experiment, subjects participated in a preference and then a grammaticality classification session. During preference classification, subjects were presented with new sequences, which they have not seen before. They were instructed to classify the new sequences according to their immediate intuitive preference (i.e. guessing whether they liked the sequence, or not, based on gut-feeling; *preference instruction*). Subsequently, they were informed about the existence of a generating set of rules and the subjects were asked to classify new sequences as grammatical or not based on their gut-feeling (*grammaticality instruction*). fMRI data were acquired during both preference and grammaticality classification (Petersson *et al.* 2010, Folia *et al.* 2011).

The classification sequences were presented via an LCD-projector on semi-transparent screen that the subject comfortably viewed through a mirror mounted on the head-coil. The classification sessions were split in two parts, in order to balance response finger within subjects (subjects indicated their decision by pushing the corresponding response key with their left/right index finger). Each part lasted approximately 20 minutes. After a 1 s pre-stimulus period, the sequences were presented sequentially, followed by a 3 s response window. A low-level baseline condition was also included; a sensorimotor decision task in which sequences of letters P or L (matched for sequence length to the classification set) were presented in the same fashion as the classification sequences and subjects responded by pressing the right or left index finger, respectively. The different sequence types were presented in random order.

### 3. Data Acquisition and Statistical Analysis

Behavioral data were analyzed with repeated measures ANOVAs (SPSS 15.0) with non-sphericity correction. A significance level of  $P < .05$  was used throughout. Data analysis was carried out for the whole group and the sub-sample for which CNTNAP2 (SNP RS7794745) data were available (T-group: AT/TA/TT allele; nonT-group: AA allele).

#### 3.1. MR Data Acquisition

Whole head T2\*-weighted functional echo planar blood oxygenation level dependent (EPI-BOLD) fMRI data were acquired with a Siemens Avanto 1.5T scanner using an ascending slice acquisition sequence (volume TR = 2.6s, TE = 40 ms, 90 degree flip-angle, 33 axial slices, slice-matrix size = 64x64, slice thickness = 3 mm, slice gap = .5 mm, FOV = 224 mm, isotropic voxel size = 3.5x3.5x3.5 mm<sup>3</sup>) in a randomized event related fashion. For the structural MR image volume, a high-resolution T1-weighted magnetization-prepared rapid gradient-echo pulse

sequence was used (MP-RAGE; volume TR = 2250 ms, TE = 3.93 ms, 15 degree flip-angle, 176 axial slices, slice-matrix size = 256x256, slice thickness = 1 mm, field of view = 256 mm, isotropic voxel-size = 1.0x1.0x1.0 mm<sup>3</sup>).

### 3.2. MR Image Pre-Processing and Statistical analysis

We used the SPM5 software for image pre-processing and statistical analysis. The EPI-BOLD volumes were re-aligned to correct for individual subject movement and were corrected for differences in slice acquisition time. The subject-mean EPI-BOLD images were subsequently spatially normalized to the functional EPI template provided by SPM5. The normalization transformations were generated from the subject-mean EPI-BOLD volumes and applied to the corresponding functional volumes. The functional EPI-BOLD volumes were transformed into the MNI space, an approximate Talairach space (Talairach & Tournoux 1988), defined by the SPM5 template, and spatially filtered with an isotropic 3D spatial Gaussian kernel (FWHM = 10 mm). The fMRI data were analyzed statistically, using the general linear model framework and statistical parametric mapping in a two-step random-effects summary-statistics procedure (Friston *et al.* 2007). We included the realignment parameters for movement artifact correction and a temporal high-pass filter (cycle cut-off at 128 s), to account for various low-frequency effects.

At the first-level, single-subject analyses were conducted. The linear model included explanatory regressors modeling the sequence presentation period from the position of the anomaly in the HNG and LNG conditions and their correct counterparts in the HG and LG conditions. This was done separately for correct and incorrect responses. The initial part of the sequences was modeled separately, as was the baseline and the inter-sequence-interval. The explanatory variables were temporally convolved with the canonical hemodynamic response function provided by SPM5. At the second-level, we generated single-subject contrast images for the correctly classified HG, LG, HNG, and LNG sequences, relative to the sensorimotor decision baseline. These were analyzed in a random-effects repeated-measure ANOVA with non-sphericity correction for repeated measures and unequal variance between conditions. Statistical inference was based on the cluster-size test-statistic from the relevant second-level SPM[T] maps thresholded at  $P = .005$  (uncorrected). Only clusters significant at  $P_{\text{FWE}} < .05$  family-wise error (FWE) corrected for multiple non-independent comparisons, based on smooth random field theory (Adler 1981, Worsley *et al.* 1996, Adler & Taylor 2007, Friston *et al.* 2007) are described. In addition, we list the coordinates of local maxima and their corresponding P-values corrected for the false discovery rate (Genovese *et al.* 2002) for descriptive purposes.

## 4. Results

### 4.1. Behavioural Results

Here we start by giving a brief summary of the most important behavioral results

for the whole group reported in Folia *et al.* (2008) and then focus on the specifics for the sub-sample for which CNTNAP2 (SNP RS7794745) data were available. As in previous studies (Forkstam *et al.* 2008), the classification performance of the whole group was well above chance for both instruction types (preference classification:  $P < .001$ ; grammaticality classification:  $P < .001$ ). Standard signal detection analysis showed a robust  $d'$ -prime effect in discriminating between G and NG sequences (preference:  $P < .001$ ; grammaticality:  $P < .001$ ). No significant response bias was found (preference and grammaticality classification  $P > .6$ ). Participants did not discriminate between high and low ACS sequences (preference:  $P > .22$ ; grammaticality:  $P > .66$ ), and there was no significant response bias (preference  $P > .98$ ; grammaticality:  $P > .8$ ).

We then analyzed the performance data in terms of endorsement rate (i.e. item classified as grammatical independent of their actual grammaticality status). In other words, if the subjects acquire significant aspects of the grammar, then they should endorse grammatical items more often than non-grammatical items. Both grammaticality status and local subsequence familiarity influenced the endorsement rate. The endorsement rate was significantly affected by grammaticality status (preference:  $P < .001$ ; grammaticality:  $P < .001$ ), and by local subsequence familiarity (preference:  $P < .001$ ; grammaticality:  $P < .001$ ), while the interaction between grammaticality status and local subsequence familiarity was non-significant (preference:  $P = .06$ ; grammaticality:  $P = .11$ ). These results show that grammaticality status is used for structural generalization in classifying novel sequences and thus provide support for the notion that grammatical structure instead of subsequence, or fragment features, determine classification (Folia *et al.* 2008).

The critical measure in the behavioral results was the preference of the participants for grammatical, and relative aversion of non-grammatical, sequences. The participants only need to indicate whether they like or dislike a given sequence and therefore we do not need to inform them about the presence of a complex rule system before classification (or at any other point of the experiment), which is the case in standard versions of the AGL paradigm, which uses grammaticality instead of preference classification. Therefore, from the subject's point of view, there is no such thing as a correct or incorrect response and the motivation to use explicit strategies is thus minimized. The participants were also strongly encouraged to trust their gut-feeling in making their decisions. Consistent with this, the subjective reports from the structured post-experimental interview showed that the participants did not utilize an explicit strategy but that their classification decisions were based on gut-feeling. Moreover, the subjective ratings of perceived performance did not correlate with the actual classification performance (Folia *et al.* 2008).

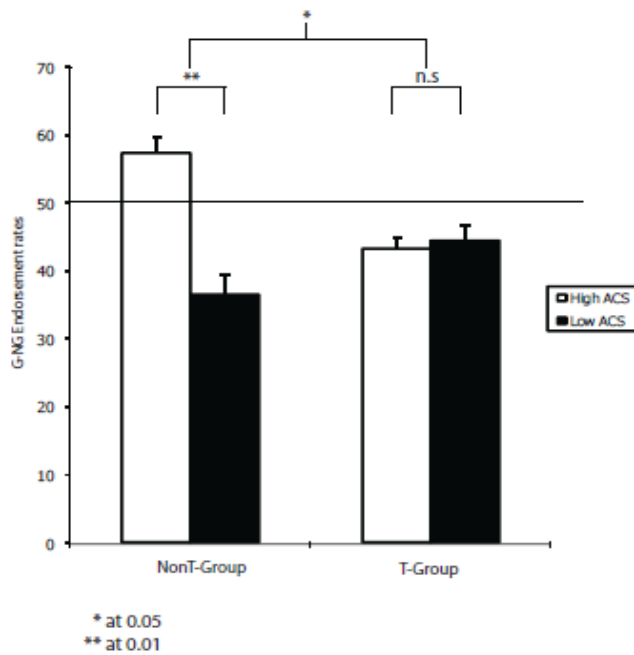


Figure 1: Grammaticality classification and CNTNAP2. The endorsement rates for grammatical and non-grammatical sequences in the T- and nonT-group. The interaction between grammaticality status and local subsequence familiarity was significant for the nonT-group (AA carriers) and not the T-group. The nonT-group thus shows greater dependence on local subsequence familiarity in making the grammaticality judgments than the T-group, despite the fact that local subsequence familiarity is not predictive for grammaticality status. Error bars corresponds to standard error of the mean.

Overall, the sub-sample for which CNTNAP2 data were available was found to behave essentially identical to the whole group and here we focus on their grammaticality classification performance. On the last day, the correct classification performance was well above chance on grammaticality classification ( $78 \pm 19\%$  correct,  $T(11) = 5.36$ ,  $P < .001$ ). Both grammaticality status and local subsequence familiarity influenced the endorsement rate. Repeated measures ANOVA showed significant main effects of grammaticality status ( $F(1,11) = 13.2$ ,  $P = .004$ ) and local subsequence familiarity ( $F(1,11) = 21.0$ ,  $P = .001$ ). We then analyzed the data with a repeated measure ANOVA with grammaticality status and local subsequence familiarity (ACS) as within-subject variables and allele (T/nonT) as between factors. Post-hoc analysis was conducted where relevant. The correct classification performance was significantly greater than chance in both groups (T-group:  $T(7) = 3.34$ ,  $P = .01$ ; nonT-group:  $T(3) = 8.25$ ,  $P = .004$ ). For grammaticality classification, the three-way interaction between grammaticality status, local subsequence familiarity, and allele group was significant ( $F(1,10) = 4.86$ ,  $P < .05$ ) as well as the main effect of grammaticality status ( $F(1,10) = 20.5$ ,  $P = .001$ ) and local subsequence familiarity ( $F(1,10) = 23.4$ ,  $P = .001$ ). No other interaction reached significance. Post-hoc analysis in the nonT-group revealed a main effect of grammaticality status ( $F(1,3) = 17.5$ ,  $P = .02$ ), and a significant interaction between grammaticality status and local subsequence familiarity ( $F(1,3) = 22.6$ ,  $P = .01$ ). In the T-group, a significant main effect was found for both grammaticality status ( $F(1,7) = 11.66$ ,  $P = .01$ ) and local

subsequence familiarity ( $F(1,7) = 17.9, P = .004$ ), while no interaction was significant (Figure 1).

These results show that the T- and the nonT-group behave similarly to the whole sample, including the development of a preference for grammaticality (Folia *et al.* 2008, Forkstam *et al.* 2008). However, the grammaticality classification performance of the T-group was independent of local subsequence familiarity (Figure 1), while this was not the case for the nonT-group. Thus, the absence of a T nucleotide in the CNTNAP2 SNP RS7794745 might be associated with a greater reliance on local subsequence familiarity (ACS) during classification. This, despite the fact that the grammaticality status is independent of local subsequence familiarity, by the construction of the stimulus material, and therefore ACS has little, if any, predictive value with respect to grammaticality status.

#### 4.2. fMRI Results

Here we briefly summarize the results reported in Folia *et al.* (2011). Preference classification compared to the sensorimotor decision baseline (Figure 2) activated a set of brain regions (cluster  $P_{FWE} < .001$ ) very similar to what has been observed in previous studies of grammaticality classification (Petersson *et al.* 2004, 2010, Forkstam *et al.* 2006). These activations included the inferior and middle frontal regions bilaterally (BA 44/45), extending into surrounding cortical regions, frontal operculum, and the anterior insula. Additional prefrontal activations included the anterior cingulate and surrounding cortex. Bilateral posterior activations included the inferior parietal cortex (BA 39/40, extending into the posterior superior temporal (BA 22), bilaterally. Bilateral occipital activations were centered on the middle and inferior occipital gyri and extended into the fusiform and the posterior mid-inferior temporal regions, as well as the cerebellum. Significant activations were also observed in the basal ganglia bilaterally, including the caudate nucleus, globus pallidus, and putamen. The results were similar for 'correctly' preferred HG- and LG sequences (Figure 2). Large, and highly significant, deactivations were found in the bilateral medial temporal lobe memory system, including the hippocampus proper (cluster  $P_{FWE} < .001$ ), replicating previous results for grammaticality classification (Petersson *et al.* 2010).

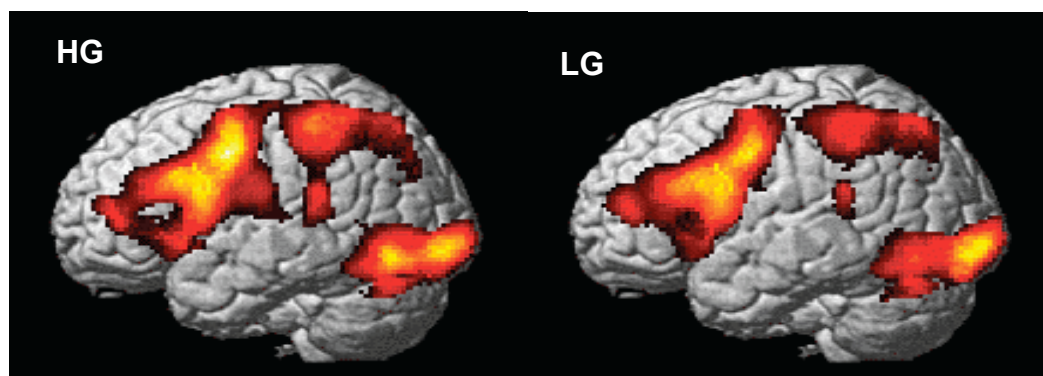


Figure 2: Preference classification. Brain regions engaged during 'correct' preference classification of grammatical sequences with high (HG) and low (LG) subsequence familiarity (ACS) relative the sensorimotor decision baseline. Adapted from Folia *et al.* (2011).

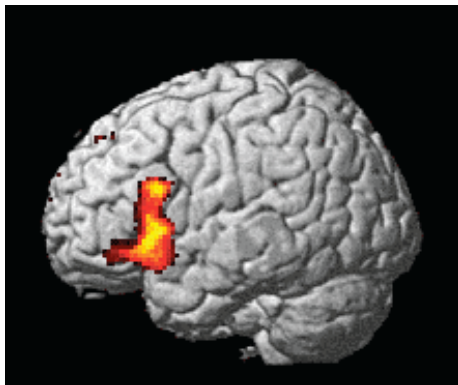


Figure 3: Preference classification. Brain regions engaged by artificial syntactic anomalies (NG > G). Adapted from Folia *et al.* (2011).

In preference classification (Folia *et al.* 2011), as in previous studies of grammaticality classification (Petersson *et al.* 2004, 2010, Forkstam *et al.* 2006), artificial syntactic anomalies (NG > G; Figure 3) engaged a network of brain regions, including the left inferior and right inferior-middle frontal gyrus (left and right cluster  $P_{\text{FWE}} < .001$ ) centered on Broca's region (BA 44/45). In the reverse contrast (G > NG), we observed no significant differences. There was no significant effect of local subsequence familiarity (cluster  $P_{\text{FWE}} > .98$ ), neither were there any significant interaction (cluster  $P_{\text{FWE}} > .83$ ), consistent with our behavioral findings (Folia *et al.* 2008).

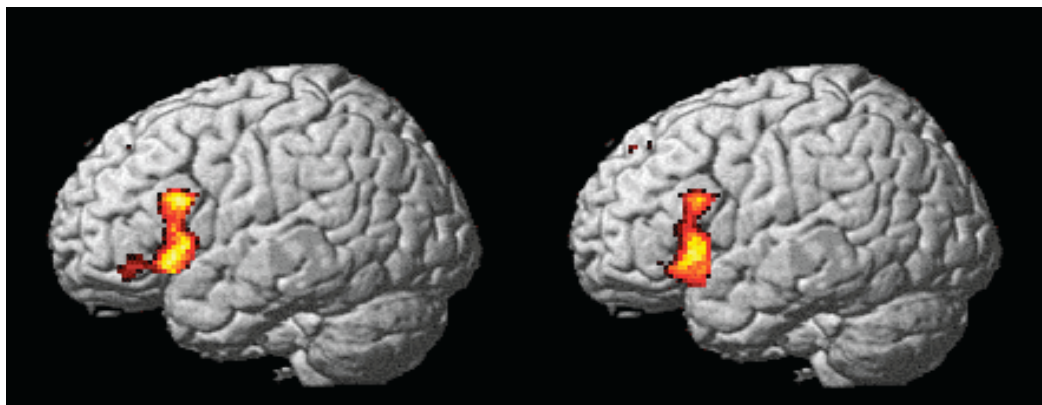


Figure 4: Brain regions engaged during both preference and grammaticality classification. Left: The NG > G effect of Folia *et al.* (2011) masked with the related effect observed in Petersson *et al.* (2010). Right: The overlap of the NG > G effect in preference classification (Folia *et al.* 2011) masked with natural syntax related variability in the same subjects observed in (Folia *et al.* 2009).

Here, we examined the overlap between preference and grammaticality classification by masking the preference classification contrast (NG vs. G effect) from Folia *et al.* (2011) with the same contrast of grammaticality classification from Petersson *et al.* (2010; Figure 4 and Appendix B). We found a common overlap in the inferior frontal regions, centered on Broca's region (BA 44/45) and extending into the frontal operculum/anterior insula, bilaterally, as well as the right middle frontal region (LIFG cluster  $P_{\text{FWE}} = .003$ ; RI/MFG cluster  $P_{\text{FWE}} < .001$ ). In addition, the anterior cingulate/supplementary motor regions were found to

be active in both the tasks (ACC/SMA cluster  $P_{\text{FWE}} = .001$ ; see Appendix B for details). Reversing the order of masking yielded identical results (LIFG cluster:  $P_{\text{FWE}} = .003$ ; RI/MFG cluster:  $P_{\text{FWE}} < .001$ ; ACC/SMA cluster  $P_{\text{FWE}} = .001$ ). Moreover, there was no significant difference between preference and grammaticality classification in any contrast, including the main effects of grammaticality status and local subsequence familiarity. Thus, artificial syntax processing engaged the same brain regions during preference and grammaticality classification, although there was a tendency that grammaticality classification yielded somewhat more robust results, highly consistent with the behavioral results (Folia *et al.* 2008, see also Forkstam *et al.* 2008). The same conclusion is reached when we examined the common overlap between artificial and natural syntax processing by masking the NG vs. G effect observed for preference classification with the natural-syntax-related variability in the same subjects (Figure 4; LIFG cluster:  $P_{\text{FWE}} = .001$ ; RI/MFG cluster:  $P_{\text{FWE}} = .008$ ; ACC/SMA cluster  $P_{\text{FWE}} = .001$ ), that is, the main effect of syntax in the 2x2 natural language experiment of Folia *et al.* (2009).

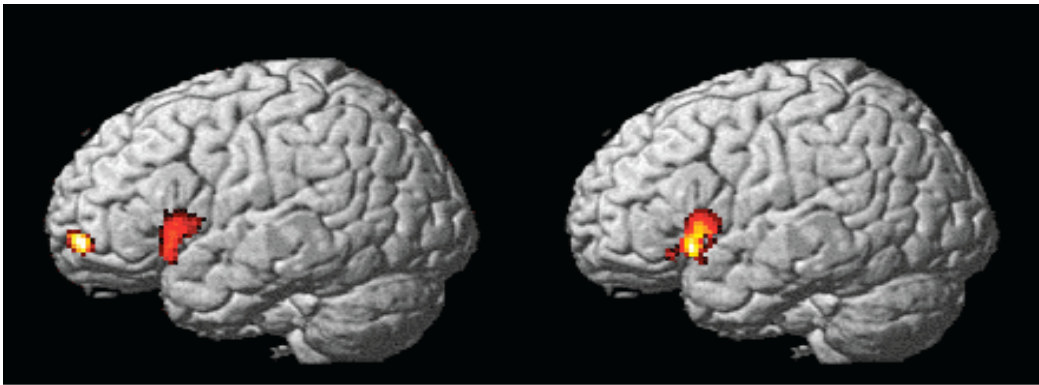


Figure 5: Brain regions differentiating the T- and the nonT-groups. Left: Group differences related to grammaticality classification (*nonT* > T). Right: Group differences related to grammatical sequences of high local subsequence familiarity (*nonT* > T).

Finally, we explored the fMRI results of Petersson *et al.* (2010; Figure 4) with respect to differences between the T- and nonT-group. The results showed significantly greater activity for the nonT-group compared to the T-group in the left inferior frontal gyrus (BA 44/45,  $P_{\text{FWE}} = .002$ ), the left fronto-polar region (BA 10,  $P_{\text{FWE}} = .012$ ), and the left ventral occipito-temporal region (BA 37,  $P_{\text{FWE}} = .003$ ) during grammaticality classification. The group difference found in Broca's region was mainly related to differences between the T- and nonT-group when processing grammatical sequences, in particular grammatical sequences of high local subsequence familiarity (BA 44/45 centered on  $[-48, 16, -2]$ ,  $P_{\text{FWE}} = .024$ ; Figure 5). The results were almost identical for the preference classification data of Folia *et al.* (2011).

## 5. Discussion

One of the main objectives of this study was to compare the brain networks engaged by preference classification and the standard grammaticality classification

task after implicit artificial syntax acquisition. The results show that preference and grammaticality classification engage virtually identical brain regions, consistent with previously reported behavioral findings (Folia *et al.* 2008, Forkstam *et al.* 2008). The theoretical advantage of preference compared to grammaticality classification is that there is no correct or incorrect response from the perspective of the participant and at no point is there a need to inform the participant about the existence of an underlying generative grammar, as is the case of the standard grammaticality classification. Nevertheless, the results show that preference and grammaticality classification are (qualitatively) equivalent both at the behavioral and brain levels. In particular, Broca's region, the left inferior frontal gyrus centred on BA 44/45, is active during the artificial syntax processing of well-formed (grammatical) sequence independent of local subsequence familiarity. Moreover, this region is engaged to a greater extent when a syntactic anomaly is present and the unification of structural treelets becomes difficult or impossible. The behavioral results of Folia *et al.* (2008) show that subjects implicitly acquired significant knowledge from being exposed to only grammatical examples and without receiving performance feedback at any stage of the experiment. Moreover, the behavioral results show that participants apply implicitly acquired structural knowledge (independent of subsequence familiarity) and the corresponding fMRI results show that brain regions central to natural syntax processing are engaged (Folia *et al.* 2011), also when they are not explicitly instructed or receives any information concerning the existence of a generative grammar. The results of this study show that the participants do so at levels comparable to grammaticality classification. Thus, the structural mere-exposure effect is a robust phenomenon at the behavioral (Folia *et al.* 2008, Forkstam *et al.* 2008) and brain level (Folia *et al.* 2011). In other words, the effects related to artificial syntax processing in the left inferior frontal region (BA 44/45) were essentially identical when we masked these with activity related to grammatical classification in the same subjects, as well as when masked with activity related to natural syntax processing in the same participants. Our results are also highly consistent with functional localization of natural language syntax in the left inferior frontal gyrus (Bookheimer 2002, Petersson *et al.* 2004, Hagoort 2005).

We used a simple right-linear unification grammar with a finite vocabulary of terminal symbols and a finite lexicon of primitive trees (*treelets*, i.e. structured lexical items; see materials and methods section for details). From an abstract point of view, unification (Vosse & Kempen 2000) is a way to implement computational control in lexicalist grammars (Forkstam & Petersson 2005, Petersson *et al.* 2005). More specifically, for a given lexical item of the grammar used in this study, for example  $[s_v [T, s_k]]$ , the features  $s_v, s_k$  can be interpreted as control features and T as a surface feature. Here, two lexical items,  $[s_v [R, s_j]]$  and  $[s_k [Q, s_l]]$ , unify (i.e. combine or merge) through a *unification operation* U if and only if  $s_j = s_k$  or  $s_l = s_v$ , a process which is incremental and recursive. For example, if the structure  $[s_1 [M, [s_2 [S, s_2]]]]$  is already present in the unification space when the lexical item  $[s_2 [V, s_4]]$  is retrieved, a larger combinatorial structure can be formed by unification  $U([s_1 [M, [s_2 [S, s_2]]]], [s_2 [V, s_4]]) = [s_1 [M, [s_2 [S, [s_2 [V, s_4]]]]]$ , and so on. We note that the control features have acquired a particular functional role in this picture, which can be described in terms of governing the unification



process based on selecting the structural arrangement that can be integrated. In a certain sense therefore, the finite-state control has been distributed over the lexicon among the lexical items in terms of control features. In essence, this retraces a major trend in theoretical linguistics in which more of the grammar is shifted into the lexicon and the distinction between lexical items and grammatical rules is beginning to vanish (cf. Joshi & Schabes 1997, Vosse & Kempen 2000, Jackendoff 2002, 2007). In this context, Broca's region can be considered as a brain region that gradually controls the outcome of parsing or generation. A related, but different proposal has recently been put forward by Bornkessel-Schlesewsky *et al.* (2010), who argue that the left inferior frontal region, including Broca's region, can be described as a brain region that controls the outcome of different processes from general to specific along the anterior-posterior direction. Bornkessel-Schlesewsky *et al.* note that their proposal is partly compatible with Hagoort's (2005) assumption of a unification gradient within the left inferior frontal gyrus.

### 5.1. A Genetic Basis for Implicit Acquisition of Structured Sequence Knowledge

Two facts about language learning seem indisputable: (i) only humans acquire language, no other species, and thus there must be some biological element that accounts for this ability; (ii) it is also clear that no matter how much of a head start the learner gains through innate constraints, language is learned. Both innate endowment and learning contribute to language acquisition, the result of which is a complex and sophisticated body of linguistic knowledge (Chomsky 1963, Chomsky & Miller 1963). It is clear that unless restrictions are placed on the available "space of possible languages" (i.e. the model space) and/or the characteristics of the acquisition mechanism (i.e. the learning dynamics), "learning" would simply reduce to storing experience (Petersson 2005a, Folia *et al.* 2010). Much of the current discussion of language acquisition concerning the nature of innate constraints is focused on whether these are linguistically specific or not (e.g. Chomsky 1986, 2005; however, see Nowak *et al.* 2002, Chomsky 2007, Christiansen & Chater 2008, Hornstein 2009). We think this is an empirical issue — however, what is clear is that no interesting, complex form of learning is possible without constraints (Vapnik 1998, Jain *et al.* 1999). In this context, Yang (2004) cites an interesting insight by Jerry Fodor (2001: 107–108), "Chomsky can with perfect coherence claim that innate, domain specific [constraints] mediate language acquisition, while remaining entirely agnostic about the domain specificity of language acquisition mechanisms". What can this possibly mean? Folia *et al.* (2010) outline several possibilities. For instance, the learning/developmental dynamics might be domain-general in form, but in the context of language acquisition, operate on a model space that is restricted by innate, language-specific constraints. By language-specific constraints we mean constraints which play no role in cognition outside the language faculty. No one doubts the existence of innate constraints, rather the issue is whether the innate constraints are specific to language or not. In fact, Folia *et al.* argue that in order to rule out innate, language-specific constraints completely, it is necessary to establish that none of the following candidates carry such constraints: (1) the initial state of the

learner; (2) the model space; (3) the learning/developmental dynamics; (4) the representational space; or (5) the representational dynamics — a difficult empirical task. Alternatively, if sufficient non-language-specific constraints for language acquisition are discovered, the necessity of language-specific constraints recedes.

In this fMRI study we took advantage of the fact that a subsample of our participants (Petersson *et al.* 2010, Folia *et al.* 2011) was part of the BIG project at the Donders Centre for Cognitive Neuroimaging and the Department of Human Genetics of the Radboud University Nijmegen. This allowed us to explore the potential role of the CNTNAP2 gene in artificial syntax acquisition at both the behavioural and the brain level. This small scale investigation of possible CNTNAP2 related effects (more precisely, effects related to the common polymorphism observed at the single nucleotide polymorphism RS7794745) in the context of artificial syntax acquisition and structured sequence processing suggests that the T-group (AT- and TT-carriers) was sensitive to the grammaticality status of the sequences independent of local subsequence familiarity. This might mean that individuals with this genotype acquire structural knowledge more rapidly, utilize the acquired knowledge more effectively, or are better able to ignore cues related to local subsequence familiarity in comparison to the nonT-group (AA carriers). This suggests differences in the implicit acquisition process between the two groups. Another possibility is that, if the two groups eventually achieve the same level of successful overall classification at the end acquisition, the nature of sequence processing might be different, since only the nonT-group is sensitive to local subsequence familiarity (which is not predictive of the grammaticality status). In contrast, the T-group relies only (or at least to a greater extent) on their implicitly acquired structural knowledge, which they successfully generalize to novel items. This suggests a qualitative, rather than a quantitative, processing difference between groups. Parallel to these behavioral findings, we observed significantly greater activation in Broca's region centered on the left BA 44/45 as well as the left frontopolar region (BA 10) in the nonT- compared to the T-group. The meaning of these fMRI differences between the two groups is unclear and requires further research for a full understanding. Nevertheless, these initial efforts suggest that it is worthwhile to investigate the genetic basis of the capacity for structured sequence processing in large-scale studies by investigating the relevant biological pathway(s) (Konopka & Geschwind 2010, Newbury & Monaco 2010, Pezawas & Meyer-Lindenberg 2010, and references therein). However, given that CNTNAP2 has been linked to specific language impairment (SLI) and provides a mechanistic link between clinically distinct syndromes involving disrupted language (Vernes *et al.* 2008), and assuming that the structured sequence learning mechanism investigated by artificial grammar learning is shared between artificial and natural syntax acquisition, the present behavioral and fMRI results might suggest that the FOXP2–CNTNAP2 pathway is somehow related to the acquisition of structured sequence knowledge as well as individual differences in artificial and natural syntax acquisition.

## 5.2. *Language as a Neurobiological System and Bounded Recursion*

Cognitive neuroscience approaches the brain as a computational system — a sys-

tem conceptualized in terms of information processing. This entails the idea that a subclass of its physical states is viewed as representations and that transitions between states can be understood as a process implementing operations on the corresponding representational structures. It is uncontroversial that any physically realizable computational system is necessarily finite with respect to its memory organization and that it processes information with finite precision (e.g., due to the presence of internal noise or architectural imprecision; Turing 1936a, 1936b, Minsky 1967, Savage 1998, Koch 1999). We have previously indicated why this state of affairs renders the Chomsky hierarchy for classical cognitive models (i.e. Church–Turing computational models) less relevant to neurobiological systems from a neurobiological processing perspective (Petersson 2005a, 2005b, 2008, Petersson *et al.* 2010). The Chomsky hierarchy is in essence a memory hierarchy and it distinguishes between (a few) complexity classes (and corresponding grammar classes) in the context of infinite (unbounded) memory. If we view the faculty of language as a neurobiological system, given its finite storage capacity and finite precision computation, the Chomsky hierarchy is less relevant — it does not make the relevant distinctions. However, bounded versions of the different memory architectures entailed by the hierarchy might be relevant (although we think these should not be taken too seriously). For example, the unbound push-down stack is a memory architecture corresponding to the class of context-free grammars, and it is conceivable that a bounded push-down stack is used in language processing, as suggested by Levelt (1974) as one possibility. Of course, this does not imply that the Chomsky hierarchy is irrelevant for computational theory (Davis *et al.* 1994, Pullum & Scholz 2010) or competence grammars in theoretical linguistics (Chomsky 1963). However, we note that modern complexity theory, which is more closely related to processing complexity rather than the Chomsky hierarchy, makes fine-grained distinctions (Cutland 1980, Papadimitriou 1993, Savage 1998, Hopcroft *et al.* 2000, Arora & Barak 2009) and might, perhaps, be useful from a neurobiological processing perspective (although this is unclear).

With the advent of generative grammar, recursion became key to achieving discrete infinity (e.g. Chomsky 1956, 1963). Accordingly, early psycholinguistics devoted considerable effort to the study of complex recursive constructions, especially in the form context-free or more general grammars (Chomsky 1963, Levelt 1974). However, it was theoretically suggested (e.g. Chomsky 1963: 329–333, 390), and soon empirically confirmed, that unbound (i.e. infinite) recursive capacity is not realizable in human performance (~actual cognitive processing). Thus, it was found that sentences with more than two center-embeddings 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 & Cairns 1970, Larkin & Burns 1977) and comprehend (Wang 1970, Hamilton & Deese 1971, Blaubergs & Braine 1974, Hakes *et al.* 1976), and are, paradoxically, judged to be ungrammatical (Marks 1968).

Recursion is once again attracting attention as an hypothesized key feature of the language faculty, with the suggestion that unbounded recursion may be the only property of the language faculty that is both species-specific and domain-specific (Hauser *et al.* 2002). Nevertheless, in order to preserve the essen-

tial feature of the notion of discrete infinity (unbounded “human creativity”), Chomsky introduced the notion of a competence grammar, “a device that enumerates [...] an infinite class of sentences with structural descriptions” (Chomsky 1963: 329–330, device A in Fig. 1). The competence grammar is distinct from the language acquisition and processing (“performance”) system (Chomsky 1963: 329–330, devices C and B, respectively, in Fig. 1). One consequence of grammars or computational models that support unbounded recursion (and infinite precision processing), is that they overgeneralize, by generating arbitrarily long sequences (and correspondingly complex sequence structures) that are never used, and in fact, has never been observed. This might or might not be a problem, depending on ones perspective on these issues. However, this is not a problem for bounded recursive procedures (or equivalent analogues, Petersson 2005b, 2008). As previously noted, one uncontroversial limitation on actual neurobiological systems is their finiteness, both in terms of memory and processing precision. For instance, Chomsky remarks that both language processing and language acquisition, “which represents actual performance, must necessarily be strictly finite”, that is, a finite-state machine (Chomsky 1963: 331–333); and continues: “Nevertheless, the performance of the speaker or hearer must be representable by a finite automaton of some sort” (p. 390). However, he further argued that “any interesting realization of B [i.e. a finite-state processing system] that is not completely ad hoc will incorporate A [i.e. a competence grammar] as a fundamental component”. One example of this idea is a (e.g., universal) Turing machine with finite tape-memory (Petersson *et al.* 2010: fn. 3). Another example is a (e.g., universal) register machine with a finite number of registers (Petersson 2005b). In both cases, it could be argued that the finite-state control unit, in a certain sense, represents unbounded ‘knowledge’ (or competence grammar) as well as unbounded recursive potential. However, this knowledge cannot be fully expressed, and the recursive potential not fully realized, because of memory limitations. But it could be argued, as Chomsky (1963) does, that if we imagine that hardware constraints can be disregarded (abstracted away), then the system instantiates the equivalent of a competence grammar, and thus unbounded ‘knowledge’, in this sense. Perhaps one way to interpret this idea, when applied to the language faculty, is in analogy with frictionless mechanics in physics — it retains instrumental value, but is not a correct description of the underlying reality (e.g., a correct model of friction is an atomic, mainly electromagnetic phenomenon).

Finite-state and finite-precision computation devices, including real neural networks, are sufficient to handle bounded recursion of general type, so there is no real problem here from the point of view of language processing (‘performance’). We think this opens the possibility for lateral thinking on matters related to the knowledge of language (‘competence’). We argue that more realistic neural models provide natural bounds on memory and on processing as well as architectural precision, and therefore, on the specification of the language faculty viewed as a neurobiological system (cf. Petersson *et al.* 2010). Generally, analog dynamical systems provide a non-classical information processing alternative to classical computational architectures (Siegelmann & Fishman 1998). In particular, network approaches offer possibilities to model cognition within a

non-classical dynamical systems framework that is natural from a neurobiological perspective. It is known theoretically, that under the assumption of infinite precision processing, Church-Turing computable processes can be embedded in dynamical systems instantiated by neural networks (e.g. Siegelmann 1999). For example, the discrete-time recurrent network can be viewed as a simple network analogue of the finite-state architecture (Petersson 2005b, Petersson *et al.* 2005). In general, the recurrent neural network architecture can be viewed as an architecture with a finite number of dynamic, analog registers (e.g., the “membrane potential”) that processes information interactively. In the simplest case, computations are determined by the network topology and by the transfer functions of the processing units, as well as the set of dynamical variables associated with these processing units. Moreover, important aspects of both short-term and long-term memory are co-localized with processing infrastructure (Petersson 2005a, Petersson *et al.* 2009). From a neurobiological perspective, therefore, it seems natural to try to understand language acquisition and language processing in terms of adaptive dynamical systems (Petersson 2005a, Petersson *et al.* 2009, 2010). Thus, an important challenge in the neurobiology of syntax is to understand syntax processing in terms of noisy spiking network processors. Similar, independent, accounts have been put forward by Culicover & Nowak (2003) in their Dynamical Grammar as well as others (Christiansen & Chater 1999, Rodriguez *et al.* 1999, Rodriguez 2001,).

What are the implications of this for theoretical models of language and grammar? The Chomsky hierarchy only has theoretical meaning in the context of infinite memory resources. Rather than giving unbounded recursion the centre stage, some of the important issues in the neurobiology of syntax, and language more generally, are related to the nature of the neural code (i.e. representation), the character of human on-line processing memory, and noisy neural finite precision computation (Koch 1999, Trappenberg 2010). Recurrent connectivity is a generic feature of brain network topology (Nieuwenhuys *et al.* 1988). Thus, recursive processing is a latent capacity in almost any neurobiological system and it would be surprising, indeed, if this feature would be unique to the faculty of language. We noted that one relevant issue from the point of view of natural language is the human capacity to process patterns of non-adjacent dependencies — not arbitrarily ‘long’ non-adjacent dependencies — there is a definite natural upper-bound set by the brain and its underlying neurophysiology. We can thus choose to work with any fruitful formal syntax framework as long as this serves its purpose, for example, to capture the presence of bounded relational patterns between lexical items in compositionally constructed sentences, to elaborate parameterized model of language acquisition or, if we are not interested in hardware constraints and implementation issues, abstract away the implementation level and explore ‘frictionless’ models of the language faculty.

## 6. Conclusion

One of the objectives of this study was to compare the brain networks engaged by artificial syntax processing during preference and grammaticality classifi-

cation after implicit artificial syntax acquisition. The results show that preference and grammaticality classification engage virtually identical brain regions, consistent with previously reported behavioral findings. In particular, the left inferior frontal region centered on BA 44/45 (Broca's region) is active during artificial syntax processing of well-formed sequences independent of local subsequence familiarity. The effects related to artificial syntax in the left inferior frontal region (BA 44/45) were essentially identical when masked with activity related to natural syntax obtained in the same subjects. Thus, the current fMRI results show that artificial syntax processing engages brain regions central to natural syntax processing. We suggest, therefore, that the left inferior frontal region is a generic on-line sequence processor that unifies information from various sources in an incremental and recursive manner. Finally, we explored CNTNAP2 related effects in artificial syntax acquisition and structured sequence processing. The results suggest that AT- and TT-carriers (at the CNTNAP2 SNP RS7794745) were sensitive to the grammaticality status independent of local subsequence familiarity, while AA-carriers were sensitive to local subsequence familiarity. We observed significantly greater activation in Broca's region and the left frontopolar region (BA 10) in the AA-carriers compared to AT- and TT-carriers. The meaning of these behavioural and fMRI findings is unclear and requires further investigation. Nevertheless, these initial efforts suggest that it is worthwhile to try to understand the genetic basis for language as well as the capacity for structured sequence processing in large-scale studies by investigating the relevant biological pathway(s).

### Appendix A: Example stimuli used for preference and grammaticality classification

Stimulus Categories	Classification Items
<i>High Grammatical (HG)</i>	VXVRXSVS
	MSSSVRXXSV
	VXSSVRXVRXSV
	MVRXSSSSVS
<i>Low Grammatical (LG)</i>	VXSVS
	MSSSSSV
	VXSVRXXXX
	MSSSVRXXXXRM
<i>High Non-Grammatical (HNG)</i>	VRVRXSSS
	MRXSSV
	VRXRXSVRXRM
	MVXSVRXXVRXRM
<i>Low Non-Grammatical (LNG)</i>	VRXRXRM
	VXVRXVXRM
	MSVRXSXRRM
	MSSVRSSSVS

**Appendix B: Overlap between preference and grammaticality classification**

<i>Anatomical region</i>	<i>Brodmann's area</i>	<i>[x y z]</i>	<i>Z-score</i>	<i>P-value</i>
<b><i>Left Inferior Frontal Cluster</i></b>				<b>.003</b>
L inferior frontal gyrus	BA 44	-54 14 2	4.07	.013
	BA 44/45	-60 20 16	3.90	.016
	BA 44/45	-52 18 22	3.81	.019
	BA 45	-60 20 10	3.54	.028
	BA 45	-46 22 22	3.35	.039
	BA 45/47	-56 18 2	4.02	.014
	BA 47	-42 20 -10	3.90	.016
L frontal operculum/ anterior insula	BA 49/13/15	-38 18 -10	4.02	.010
<b><i>Right Inferior-Middle Frontal Cluster</i></b>				<b>&lt; .001</b>
R inferior frontal gyrus	BA 44/45	50 24 18	3.65	.023
	BA 45	56 30 12	3.54	.028
	BA 45/47	58 32 0	3.42	.035
	BA 47	46 32 -4	5.08	.010
	BA 47/11	48 44 -14	3.58	.026
R mid-anterior insula	BA 13/15	40 20 -6	4.15	.011
R frontal operculum/ anterior insula	BA 49/15	36 20 -10	3.87	.017
	BA 49/13/15	32 26 0	3.57	.027
R inferior-middle frontal gyrus	BA 45/46	46 34 12	3.41	.036
	BA 45/46	48 30 16	3.39	.037
	BA 45/46	58 34 16	3.37	.038
	BA 45/46	58 34 8	3.23	.047
	BA 46	52 40 18	3.27	.044
<b><i>Medial Prefrontal-Frontopolar Cluster</i></b>				<b>.001</b>
Anterior cingulate/supplementary motor	BA 8	0 26 52	4.41	.010
	BA 8/32	6 30 44	4.60	.010
	BA 8/32	14 16 48	3.38	.037
	BA 6/8/32	-6 14 54	4.04	.013
Anterior cingulate cortex	BA 32	8 34 38	4.53	.010
	BA 32	10 32 24	4.27	.010

*Local maxima observed for correctly classified non-grammatical vs. grammatical items. Cluster P-values are family-wise error corrected and P-values of local maxima are corrected based on the false-discovery rate.*

**References**

Abrahams, Brett S., Dmitri Tentler, Julia V. Perederiy, Michael C. Oldham, Giovanni Coppola & Daniel H. Geschwind. 2007. Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proceedings of the National Academy of Sciences of the United States of America* 104, 17849–17854.

- Adler, Robert J. 1981. *The Geometry of Random Fields*. New York: Wiley and Sons.
- Adler, Robert J. & Jonathan E. Taylor. 2007. *Random Fields and Geometry*. New York: Springer.
- Alarcón, Maricela, Brett S. Abrahams, Jennifer L. Stone, Jacqueline A. Duvall, Julia V. Perederiy, Jamee M. Bomar, Jonathan Sebat, Michael Wigler, Christa L. Martin, David H. Ledbetter, Stanley F. Nelson, Rita M. Cantor & Daniel H. Geschwind. 2008. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *American Journal of Human Genetics* 82, 150–159.
- Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts & Peter Walter. 2007. *Molecular Biology of the Cell*. New York: Garland Science.
- Arora, Sanjeev & Boaz Barak. 2009. *Computational Complexity: A Modern Approach*. Cambridge: Cambridge University Press.
- Bishop, Dorothy. 2009. Genes, cognition, and communication. *Annals of the New York Academy of Sciences* 1156, 1–18.
- Blaubeergs, Maija S. & Martin D. S. Braine. 1974. Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology* 102, 745–748.
- Bookheimer, Susan. 2002. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience* 25, 151–188.
- Bornkessel-Schlesewsky, Ina, Tanja Grewe & Matthias Schlesewsky. 2010. Prominence vs. aboutness in sequencing: A functional distinction within the left inferior frontal gyrus. *Brain and Language* doi:10.1016/j.bandl.2010.06.004.
- Chomsky, Noam. 1956. Three models for the description of language. *IEEE Transactions on Information Theory* 2, 113–124.
- Chomsky, Noam. 1963. Formal properties of grammars. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, 323–418. New York: John Wiley.
- Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.
- Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.
- Chomsky, Noam. 2007. Of mind and language. *Biolinguistics* 1, 9–27.
- Chomsky, Noam & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, 269–321. New York: John Wiley.
- Christiansen, Morten H. & Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23, 157–205.
- Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31, 489–558.
- Christiansen, Morten H., Louise M. Kelly, Richard C. Shillcock & Katie Greenfield. 2010. Impaired artificial grammar learning in agrammatism. *Cognition* 116, 382–393.
- 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.



- Culicover, Peter W. & Andrzej Nowak. 2003. *Dynamical Grammar: Volume Two of Foundations of Syntax*. Oxford: Oxford University Press.
- Cutland, Nigel J. 1980. *Computability: An Introduction to Recursive Function Theory*. Cambridge: Cambridge University Press.
- Davidson, Eric H. 2006. *The Regulatory Genom: Gene Regulatory Networks in Development and Evolution*. San Diego, CA: Academic Press.
- Davidson, Eric H., Jonathan P. Rast, Paola Oliveri, Andrew Ransick, Cristina Calestani, Chiou-Hwa Yuh, Takuya Minokawa, Gabriele Amore, Veronica Hinman, Cesar Arenas-Mena, Ochan Otim, C. Titus Brown, Carolina B. Livi, Pei Yun Lee, Roger Revilla, Alistair G. Rust, Zheng jun Pan, Maria J. Schilstra, Peter J. C. Clarke, Maria I. Arnone, Lee Rowen, R. Andrew Cameron, David R. McClay, Leroy Hood & Hamid Bolouri. 2002. A genomic regulatory network for development. *Science* 295, 1669–1678.
- Davis, Martin, Ron Sigal & Elaine J. Weyuker. 1994. *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science*. San Diego, CA: Academic Press.
- Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco & Svante Pääbo. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418, 869–872.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Fodor, Jerry A. 2001. Doing without *What's Within*: Fiona Cowie's criticism of nativism. *Mind* 110, 99–148.
- Folia, Vasiliki, Christian Forkstam, Peter Hagoort & Karl Magnus Petersson. 2009. Language comprehension: The interplay between form and content. *Proceedings of the Cognitive Science Society* 2009, 1686–1691.
- Folia, Vasiliki, Christian Forkstam & Karl Magnus Petersson. 2011. Implicit structured sequence learning. Submitted.
- Folia, Vasiliki, Julia Uddén, Christian Forkstam, Martin Ingvar, Peter Hagoort & Karl Magnus Petersson. 2008. Implicit learning and dyslexia. *Annals of the New York Academy of Sciences* 1145, 132–150.
- Folia, Vasiliki, Julia Uddén, Christian Forkstam & Karl Magnus Petersson. 2010. Artificial language learning in adults and children. *Language Learning* 60(s2), 188–220.
- 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.
- Forkstam, Christian & Karl Magnus Petersson. 2005. Syntactic classification of acquired structural regularities. *Proceedings of the Cognitive Science Society* 2005, 696–701.
- 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.

- Friston, Karl J., John T. Ashburner, Stefan J. Kiebel, Thomas E. Nichols, William D. Penny (eds.). 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. San Diego, CA: Academic Press.
- Genovese, Christopher R., Nicole A. Lazar & Thomas Nichols. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878.
- Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- Gómez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4, 178–186.
- Hagoort, Peter. 2005. On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences* 9, 416–423.
- Hakes, David T., Judith S. Evans & Linda L. Brannon. 1976. Understanding sentences with relative clauses. *Memory & Cognition* 4, 283–290.
- Hakes, David T. & Helen S. Cairns. 1970. Sentence comprehension and relative pronouns. *Perception & Psychophysics* 8, 5–8.
- 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.
- Hockett, Charles F. 1963. The problem of universals in language. In Joseph H. Greenberg (ed.), *Universals of Language*, 1–29. Cambridge, MA: MIT Press.
- Hockett, Charles F. 1987. *Refurbishing Our Foundations: Elementary Linguistics from an Advanced Point of View*. Philadelphia, PA: John Benjamins.
- Hopcroft, John E., Rajeev Motwani & Jeffrey D. Ullman. 2000. *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.
- Hornstein, Norbert. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.
- Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jackendoff, Ray. 2007. A Parallel Architecture perspective on language processing. *Brain Research* 1146, 2–22.
- Jain, Sanjay, Daniel Osherson, James S. Royer & Arun Sharma. 1999. *Systems That Learn*. Cambridge, MA: MIT Press.
- Joshi, Aravind K. & Yves Schabes. 1997. Tree-adjoining grammars. In Grzegorz Rozenberg & Arto Salomaa (eds.), *Handbook of Formal Languages, vol. 3: Beyond Words*, 69–124. Berlin: Springer Verlag.
- Knowlton, Barbara J. & 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.
- Koch, Christof. 1999. *Biophysics of Computation: Information Processing in Single Neurons*. New York: Oxford University Press.
- Konopka, Genevieve, Jamee M. Bomar, Kellen Winden, Giovanni Coppola,

- Zophonias O. Jonsson, Fuying Gao, Sophia Peng, Todd M. Preuss, James A. Wohlschlegel & Daniel H. Geschwind. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462, 213–217.
- Konopka, Genevieve & Daniel H. Geschwind. 2010. Human brain evolution: Harnessing the genomics (r)evolution to link genes, cognition, and behavior. *Neuron* 68, 231–244.
- Lai, Cecilia S. L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem & Anthony P. Monaco. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523.
- Larkin, Willard & David Burns. 1977. Sentence comprehension and memory for embedded structure. *Memory & Cognition* 5, 17–22.
- Levelt, Willem J.M. 1974. *Formal Grammars in Linguistics and Psycholinguistics*, vol. III: *Psycholinguistic Applications*. The Hague: Mouton.
- MacDermot, Kay D., Elena Bonora, Nuala Sykes, Anne-Marie Coupe, Cecilia S.L. Lai, Sonja C. Vernes, Faraneh Vargha-Khadem, Fiona McKenzie, Robert L. Smith, Anthony P. Monaco & Simon E. Fisher. 2005. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *The American Journal of Human Genetics* 76, 1074–1080.
- Manza, Louis & Robert F. Bornstein. 1995. Affective discrimination and the implicit learning process. *Consciousness & Cognition* 4, 399–409.
- Marks, Lawrence E. 1968. Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior* 7, 965–967.
- Meulemans, Thierry & Martial van der Linden. 1997. Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23, 1007–1028.
- 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. 2009. Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. *Proceedings of the Cognitive Science Society* 2009, 177–182.
- 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.
- Newbury, Dianne F. & Anthony P. Monaco. 2010. Genetic advances in the study of speech and language disorders. *Neuron* 68, 309–320.
- Nieuwenhuys, Rudolf, Jan Voogd, Christiaan van Huijzen. 1988. *The Human Central Nervous System: A Synopsis and Atlas*, 3rd revised edn. Berlin: Springer Verlag.

- Nowak, Martin A., Natalia L. Komarova & Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417, 611–617.
- O'Donnell, Timothy J., Marc D. Hauser & W. Tecumseh Fitch. 2005. Using mathematical models of language experimentally. *Trends in Cognitive Sciences* 9, 284–289.
- Packard, Mark G., & Barbara J. Knowlton. 2002. Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience* 25, 563–593.
- Papadimitriou, Christos H. 1993. *Computational Complexity*. Upper Saddle River, NJ: Addison Wesley.
- Petersson, Karl Magnus. 2005a. *Learning and Memory in the Human Brain*. Stockholm: Karolinska University Press.
- Petersson, Karl Magnus. 2005b. 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, 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.
- Petersson, Karl Magnus, Christian Forkstam & Martin Ingvar. 2004. Artificial syntactic violations activate Broca's region. *Cognitive Science* 28, 383–407.
- Petersson, Karl Magnus, Peter Grenholm & Christian Forkstam. 2005. Artificial grammar learning and neural networks. *Proceedings of the Cognitive Science Society* 2005, 1726–1731.
- Petersson, Karl Magnus, Martin Ingvar & Alexandra Reis. 2009. Language and literacy from a cognitive neuroscience perspective. In David R. Olson & Nancy Torrance (eds.), *Cambridge Handbook of Literacy*, 152–182. Cambridge: Cambridge University Press.
- Pezawas, Lukas & Andreas Meyer-Lindenberg. 2010. Imaging genetics: Progressing by leaps and bounds. *NeuroImage* 53, 801–803.
- Poliak, Sebastian, Leora Gollan, Ricardo Martinez, Andrew Custer, Steven Einheber, James L. Salzer, James S. Trimmer, Peter Shrager & Elinor Peles. 1999. Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K<sup>+</sup> channels. *Neuron* 24, 1037–1047.
- 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.
- Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 5, 855–863.
- Rodriguez, Paul. 2001. Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation* 13, 2093–2118.
- Rodriguez, Paul, Janet Wiles & Jeffrey L. Elman. 1999. A recurrent neural network that learns to count. *Connection Science* 11, 5–40.
- Saffran, Jenny, Marc Hauser, Rebecca Seibel, Joshua Kapfhamer, Fritz Tsao & Fiery Cushman. 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.

- Savage, John E. 1998. *Models of Computation*. Reading, MA: Addison-Wesley.
- Siegelmann, Hava T. 1999. *Neural Networks and Analog Computation: Beyond the Turing Limit*. Basel: Birkhäuser.
- Siegelmann, Hava T. & Shmuel Fishman. 1998. Analog computation with dynamical systems. *Physica D* 120, 214–235.
- Snijders, Tom M., Mark Rijpkema, Barbara Franke, Han G. Brunner, Dan Dediu, Vasiliki Folia, Julia Uddén, Guillén Fernández, Karl Magnus Petersson & Peter Hagoort. 2011. A common CNTNAP2 polymorphism affects functional and structural aspects of language-relevant neuronal infrastructure. Submitted.
- Stadler, Michael A. & Peter A. Frensch (eds.). 1998. *Handbook of Implicit Learning*. London: SAGE.
- Talairach, Jean & Pierre Tournoux. 1988. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System — An Approach to Cerebral Imaging*. New York: Thieme Medical Publishers.
- Trappenberg, Thomas P. 2010. *Fundamentals of Computational Neuroscience* (2nd edn.). Oxford: Oxford University Press.
- Turing, Alan. 1936a. On computable numbers with an application to the Entscheidungs problem (part 1). *Proceedings of the London Mathematical Society* 42, 230–240.
- Turing, Alan. 1936b. On computable numbers with an application to the Entscheidungs problem (part 2). *Proceedings of the London Mathematical Society* 42, 241–265.
- Uddén, Julia, Susana Araujo, Christian Forkstam, Martin Ingvar, Peter Hagoort & Karl Magnus Petersson. 2009. A matter of time: Implicit acquisition of recursive sequence structures. *Proceedings of the Cognitive Science Society* 2009, 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.
- Uddén, Julia, Martin Ingvar, Peter Hagoort & Karl Magnus Petersson, K. 2011. Broca's region: A causal role in implicit processing of grammars with adjacent and non-adjacent dependencies. Submitted.
- Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92, 231–270.
- Vapnik, Vladimir. 1998. *Statistical Learning Theory*. New York: Wiley and Sons.
- Vargha-Khadem, Faraneh, David G. Gadian, Andrew Copp & Mortimer Mishkin. 2005. FOXP2 and the neuroanatomy of speech and language. *Nature Reviews Neuroscience* 6, 131–138.
- Vernes, Sonja C., Dianne F. Newbury, Brett S. Abrahams, Laura Winchester, Jérôme Nicod, Matthias Groszer, Maricela Alarcón, Peter L. Oliver, Kay E. Davies, Daniel H. Geschwind, Anthony P. Monaco & Simon E. Fisher. 2008. A functional genetic link between distinct developmental language disorders. *New England Journal of Medicine* 359, 2337–2345.
- Vosse, Theo & Gerard Kempen. 2000. Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a

- lexicalist grammar. *Cognition* 75, 105–143.
- 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.
- Wang, Marilyn D. 1970. The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior* 9, 398–404.
- Watkins, Kate E., Nina F. Dronkers, Faraneh Vargha-Khadem. 2002. Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain* 125, 452–464.
- Worsley, Keith J., Sean Marrett, Peter Neelin, Alain C. Vandal, Karl J. Friston & Alan C. Evans. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* 4, 58–73.
- Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8, 451–456.
- Zajonc, Robert B. 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement* 9, Part 2.
- Zizak, Diane M. & Arthur S. Reber. 2004. Implicit preferences: The role(s) of familiarity in the structural mere exposure effect. *Consciousness & Cognition* 13, 336–362.

Vasiliki Folia<sup>1,2,3</sup>

[vasiliki.folia@fcdonders.nl](mailto:vasiliki.folia@fcdonders.nl)

<sup>1</sup> Max Planck Institute for Psycholinguistics  
P.O. Box 310  
6500 AH Nijmegen  
The Netherlands

Christian Forkstam<sup>4</sup>

[christian@forkstam.se](mailto:christian@forkstam.se)

<sup>2</sup> Radboud University Nijmegen  
Donders Institute for Brain, Cognition and Behaviour  
P.O. Box 9101  
6500 HB Nijmegen  
The Netherlands

Martin Ingvar<sup>3</sup>

[martin.ingvar@ki.se](mailto:martin.ingvar@ki.se)

<sup>3</sup> Karolinska Institutet  
Stockholm Brain Institute  
Cognitive Neurophysiology Research Group  
17177 Stockholm  
Sweden

Peter Hagoort<sup>1,2</sup>

[peter.hagoort@mpi.nl](mailto:peter.hagoort@mpi.nl)

<sup>4</sup> Universidade do Algarve  
Institute of Biotechnology & Bioengineering CBME  
Cognitive Neuroscience Research Group  
8500-139 Faro  
Portugal

Karl Magnus Petersson<sup>1,2,4</sup>

[karl-magnus.petersson@mpi.nl](mailto:karl-magnus.petersson@mpi.nl)

# An Uncouth Approach to Language Recursivity

Eleonora Russo & Alessandro Treves

A simple-minded view is presented here on the problem of the origin of language, which dismisses any relation with hitherto unobserved specific language microcircuits in the cortex as well as with gross connectional hierarchies which are seen also in other mammals. In this view, language arises out of a capacity for spontaneous latching dynamics, which emerges when the connectivity of an extensive cortical network, which need not be hierarchical, crosses a critical phase-transition threshold.

*Keywords:* associative memory; latching dynamics; origins of language; phase transitions; Potts units; storage capacity

## 1. Cortical vs. Cognitive Organization

Neuroanatomy, proceeding slowly but surely at the tranquil pace of a descriptive science, would have the authority to inform a basic understanding of the neural mechanisms subserving higher cognitive capacities. For the faculty of language, this has not happened. There is a fundamental mismatch between the conceptual structures invoked to describe the complexity of language — parsing trees, hierarchies of grammars, principles and parameters — and those emerging from the observation of the articulation of the human nervous system. Neuroanatomical dynamics unfold over evolutionary time scales of millions of years, and their main organizational principles have been scholarly described for about a hundred years (e.g., Lorente de Nó 1938). Language dynamics, even though in the most stable parametric aspects may stretch over several thousand years (Longobardi & Guardiano 2009), unleash their astonishing power in the rapid acquisition of a language by a child — in a few years.

As a result, linguists tend to ignore taking stock of the stable organization apparent in the human brain, and at times nurture the mistaken expectation that a sudden discovery from the world of biology, like that of the structure of DNA, will at some point revolutionize the relation between language and the brain, and crack the neural codes for syntax. The Language Acquisition Device (LAD), a remarkable abstract construct (Briscoe 2000), might then acquire the semblance of a neuronal apparatus, taken to be hiding, like the Holy Grail, perhaps in one of the frontal sulci, disguised to non-believers as standard cortical circuitry. While the quest for the LAD goes on, allured by reports of quantitative differences and asymmetries in area 44 or 45 (Uylings *et al.* 2006, Amunts *et al.* 2010), it may be



useful to briefly review salient features of cortical organization.

### 1.1. *Origin and Evolution of the Cerebral Neocortex*

Higher mental processes in the human species massively involve the cerebral cortex — though, it should be noted, not to the exclusion of other structures such as the cerebellum or the basal ganglia. The cerebral cortex derives from a structure, the pallium or dorsal component of each hemisphere, as it bulges out of the fore-brain behind each olfactory bulb, that was presumably common also to the ancestors of reptiles and birds, and whose ancient amniotic phenotype is thought to resemble most closely the cortex of modern reptiles. In the evolution of mammalian lineages, the two dramatic events that separated them from other amniotes were the lamination of the dorsal cortex and the reorganization of the medial cortex into the modern mammalian hippocampus. These two events likely occurred between three and two hundred million years ago, with outcomes that are anatomically clear and functionally obscure, but in any case are very much with us to this day. No further major reorganization has occurred since, common to all mammalian species, and differences in the organization of specific radiations, such as the incomplete granulation of cetacean cortex (Huggenberger 2008), are more in the way of amendments than bright new ideas.

The two remarkably stable traits of a 3-fold differentiated hippocampus and a 3-fold laminated cortex (Fig. 1) seem to us to express the fundamental mammalian *geist*, and the crucial challenge for theories of mammalian neural computation to try and explain; yet we feel rather isolated in our interest for these two phase transitions (Treves 2003, 2004).

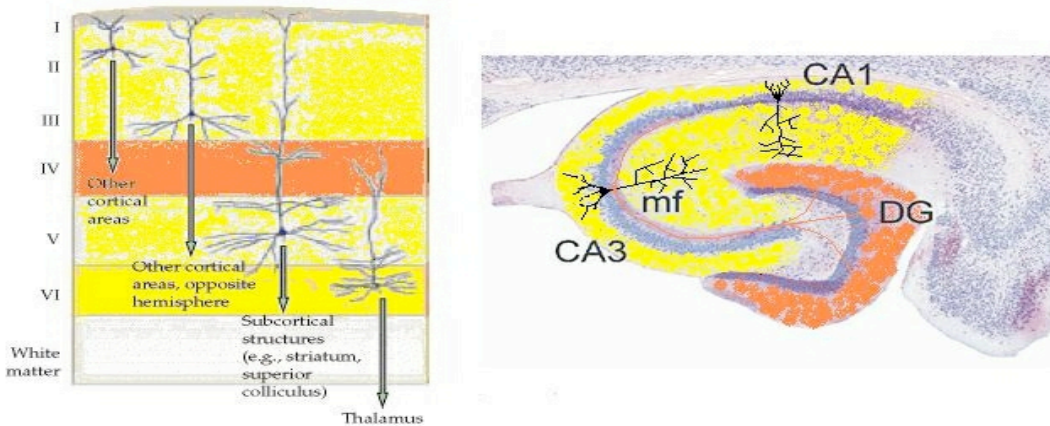


Figure 1: In mammals, the dorsal cortex is laminated (left) and the medial cortex is reorganized into the hippocampus, differentiated into 3 main sub-regions (right). In both structures the essential mammalian innovation is the insertion of an input layer of granule cells (orange) feeding into the pyramidal cells (yellow). Both these traits are common to all mammals, and neither has anything to do with the faculty of language.



Be it as it may, it is hard to identify further qualitative jumps in the structure of our nervous system, beyond these that we underwent together with other mammals. Although the simple-minded notion of cortical uniformity (Rockel *et al.* 1980) has been fiercely criticized (Herculano-Houzel *et al.* 2008, Rakic 2008), modifications and specializations on the basic mammalian design are prevailingly quantitative (Krubitzer 1995, Semendeferi *et al.* 2010). Particularly in relation to the language faculty, they come nowhere close to drawing a boundary, for example, between our brain and that of non-speaking apes, as sharp as can be drawn between, say, bats (mammals) and starlings (birds) — however startlingly recursive starlings may be (Gentner *et al.* 2006).

The dominant neuronal constituents of the cerebral cortex are the pyramidal cells, which are defined by a long axis oriented perpendicular to the cortical sheet. The main new element which sets the mammalian neocortex apart from the reptilian paleocortex is the insertion of a layer of granule cells, layer IV, where many afferent inputs terminate, and which separates infragranular from supragranular pyramidal layers (Fig. 1, left). It is thought that the new arrangement facilitates a precise point-to-point afferent connectivity, enabling the formation, in part by self-organizing processes, of fine topographical maps (Treves 2003). This occurs throughout the expanse of the so-called isocortex, although layer IV is more prominent in sensory cortices, consistent with its role in setting up orderly sensory representations, which become progressively fuzzier in more advanced areas. In frontal cortex, layer IV may remain undifferentiated, suggesting that the cognitive enhancement that accompanied, in the most advanced mammals, the quantitative expansion of the frontal lobes, is not picking up on the technological advance of a laminated cortex. In particular, the faculty of language, which arises recently and only in the human species, appears unrelated to lamination, which emerged hundreds of millions of years ago, in all mammals. It seems doubtful, therefore, to ascribe special significance to limited differences in the exact density of distinct cortical layers (Amunts *et al.* 2004). The concurrent differentiation of the medial cortex into the regions DG, CA3 and CA1 of the hippocampus (Fig. 1, right), with their strikingly idiosyncratic circuitry (Treves *et al.* 2008), is likewise common to all mammals and entirely unrelated to language.

## 1.2. *Cortical Hierarchies*

In the reptilian cortex, one distinguishes a lateral, a dorsal and a medial portion, and the olfactory origin of this most anterior component of the brain remains engraved in a clear direction of olfactory information flow, from lateral through dorsal to medial, leading to the old-fashioned but phylogenetically correct interpretation of the hippocampus as the motor division of the olfactory sensory-motor circuit. Even in those mammals, in which lateral olfactory cortex is reduced to little more than a residual, olfactory information from the bulb accesses it directly, a reminder of its former primacy. The other modalities, notably vision, audition and somatic sensation, which long ago penetrated and colonized the cortex from their more posterior original stations, send afferent inputs not directly but through the thalamus to the dorsal cortex, and there join the flow towards the medial component. In more complex mammals these new ‘settlers’ are

seen to have established an increasing number of visual, auditory and somatosensory maps, leading to the great expansion of the dorsal cortex. Such maps, including advanced processing areas in which topography has become so vague as to have been erstwhile overlooked (Kaas 1997), are laid out in a hierarchy from peripheral (primary sensory) towards central/medial. The hierarchy finds unambiguous expression in the arrangement of the cortico-cortical connections: Lower (more peripheral) areas send inputs mainly from their supragranular layers towards the layer IV (whether or not populated by granular cells) of higher areas, which send, mainly from their infragranular layers, back projections to lower areas, which terminate there in layer I. This connectivity pattern allows neuroanatomists to determine which of two areas is higher up in the hierarchy.

Originally the pallium, like the dorsal aspect of the spinal chord, is sensory. The motor cortex is thought to have differentiated relatively late, with mammals, from parts of the somatosensory cortex. Does the motor cortex, which has no layer IV, sit on top of the cortical hierarchy defined by cortico-cortical connectivity? Hardly so. It is the hippocampus, and other limbic components adjacent to it, that the laminar origin and termination of the projections elect as their supreme leaders (Barbas 1986). In fact, both in the temporal and in the frontal lobes it is the more lateral areas, which are more laminated, that project axons that terminate deep into the medial temporal or caudal orbitofrontal cortices respectively, which therefore are higher order in the connectional sense (Rempel-Clower & Barbas 2000). Information flows towards the limbic system, towards establishing memories, whereas the stream leading to motor cortex and the expression of overt behavior is more of a diverted back projection towards the periphery. It should be noted that these connectivity patterns are not rigorous rules but rather statistical trends, and that they may just reflect fossil functionalities, simply because no more modern organizational principle has come to supersede them. Even in this perspective, that is a connectional hierarchy we certainly have.

What are the changes that have then occurred in our neocortex, over this last couple of hundred million years? The most striking one, across several mammalian species, is in size, accompanied however by the parcellation of the cortical expanse into an increasing number of areas, in what is known as the process of arealization. Areas are defined by sometimes very subtle differences in cellular composition or laminar organization, that make the cortex resemble a patchwork of tonalities of the same hue, more than a blanket with continuously shaded colors. In sensory regions it appears very clearly that boundaries between areas are determined by the edges of topographic sensory representations, whether or not salient histological differences are seen with the neighboring area. In so-called simple mammals, a sensory modality may be represented by one or two topographic maps, before information is fully mixed with that from other modalities; in complex and more arealized mammals like ourselves, tens of distinct areas, each a complete map of sensory space, may represent a single modality. Several areas present further granular or quasi-granular structures, such as the mini-column, hyper-columns and pinwheels of cat and monkey visual cortex or the barrels of rat somatosensory cortex (Kaas 1997). These traits, sometimes popularized as evidence for a generic columnar organization principle, appear instead to be specializations that, in certain areas and in certain species, refine *ad*

*hoc* the common design expressed by laminated neocortex (Rakic 2008); the proliferation of distinct areas, in contrast, appears as a universal option available across species and modalities, and utilized more by some and less by others.

What keeps cortical areas together? Cortico-cortical connections. Unlike local connections, comprised by axons that never leave the gray matter or by the early departing collateral branches of those that do, cortico-cortical connections travel through the white matter, linking together areas that may sit at different levels of the periphero-hippocampal hierarchy, or at about the same level. Since cortico-cortical connections linking areas at about the same level, as well as local connections, are not hierarchical, and local and non-local connections are estimated to come in roughly equivalent numbers, it is likely that strictly hierarchical connections are a minority, unlike what is assumed in certain neural network models. The cortex is largely democratic, and a given unit can usually find itself, depending on the circumstances of life, pre-synaptic or postsynaptic to another unit, or both. Causal reasoning, which informs many conceptual models of cognitive processes, is manifestly inadequate to capture the web of potentially reciprocal influences that cortical neurons (and cortical areas) exert on one another.

Whereas a democratic arrangement of all cortical neurons on the same footing is unique, many distinct ways can be conceived of ordering them in a strict or loose hierarchical arrangement (Fuster 2009). There are, in fact, at least two more ways to define a connectional hierarchy, beyond that based on the laminar pattern of connections between areas. One is to focus on the number of synaptic contacts on the basal dendrites of pyramidal cells, that Guy Elston (2000) has estimated to increase dramatically going from occipital to temporal and then frontal cortex. Basal dendrites receive mainly local recurrent excitation, so the observation suggests a posterior-frontal gradient from more input-driven to more recurrent circuits. Another way is to focus on the density of terminals of neuro-modulators, in particular dopamine, which is particularly high in prefrontal cortex. This indicates a shift from a more rigid, operationally stable processing mode in posterior cortex to something more subject to multiple modulating influences in the front. Note that a dopamine gradient is seen also in birds, in the near absence of a full-fledged (and in any case mono-layer) cortex. These connectional/anatomical hierarchies therefore partially overlap with each other, but only in a loose sense, and there is no evidence that they are evolutionarily related, or geared towards a common functional purpose.

### 1.3. *Task Switches and Prefrontal Cortex*

The scientific study of animal behavior has been in part based on the experimental paradigm of classical conditioning, whereby, since Pavlovian times, an animal is exposed to stimulus  $x$  followed, in rapid temporal succession, by another stimulus  $y$ . In its 'operant' variant the animal learns instead that to stimulus  $x$  it must respond  $y$ , to get reward  $z$  (or to avoid a punishment). In their crude simplicity such paradigms lend themselves to quantitative measures, more than ecological behavior, and have been extended in several directions. A trivial extension is to increase the number of stimulus-response associations, for example, to  $x_1$  the subject should respond  $y_1$  to get  $z$ , whereas to  $x_2$  the subject should res-

pond  $y_2$ , to  $x_3$ ,  $y_3$  and so on. Another type of extension, also increasingly used with human subjects, involves the combinatorial articulation of contingencies, for example, in context  $w_1$ , to  $x_1$  one should respond  $y_1$ , and to  $x_2$ ,  $y_2$ , whereas in context  $w_2$ , to  $x_1$  one should respond  $y_2$ , and to  $x_2$ ,  $y_1$ . The paradigm can obviously be complexified by expanding the number of associations, but also by adding levels, for example, in task  $v_1$  when in context  $w_1$ , to  $x_1$  respond  $y_1$ , while in  $w_2$ , to  $x_1$  respond  $y_2$ , and vice versa in task  $v_2$ . Obviously, if elements  $v$ ,  $w$ ,  $x$  are purely labels, their complete configuration can be recapitulated in a compositional variable, say  $u$ , where, for example,  $u_1$  denotes 'molecular' configuration  $(v_1, w_1, x_1)$ ,  $u_2$  denotes configuration  $(v_1, w_2, x_1)$ , etc. If, however, elements  $v$ ,  $w$ ,  $x$  are presumed to have a life of their own, both in the real world and as represented in the brain of subjects, it is convenient to maintain an atomic notation, to point out that a certain stimulus–response association, for example,  $x_1 - y_1$ , is correct in task  $v_1$  and in context  $w_1$ , but not in situation  $(v_2, w_1)$ . The experimenter can add levels of contingency  $n$ ,  $n+1$ ,  $n+2$ , *ad libitum*, making the paradigm progressively more complex.

A recurring observation from the analysis of brain lesioned patients and from neuroimaging studies is that the most anterior portions of the cortex appear to be involved in the correct learning and execution of the higher contingency levels, with perhaps the frontal pole necessary for the maximum complexity level  $n_{\max}$  in a particular type of paradigm, that normal human subjects are able to deal with. While there is no evidence that  $n_{\max}$  is universal across different paradigms, the review of neuroimaging data and the *ad hoc* experiments designed by Badre & D'Esposito (2007) indicate that  $n_{\max}$  can be higher than 3, in the sense that they can distinguish at least 4 hierarchical levels of contingency processing in a series of paradigms of increasing complexity.

One may represent such paradigms as a tree, where each variable is a branch generating at the end node the various thinner branches it can lead to. For example, branch  $w_1$  'generates'  $x_1$ ,  $x_2$  e  $x_3$ , which in turn 'generate'  $y_5$  and  $y_8$ . The tree representation misses out the combinatorial nature of the process, because if a branch represents  $x_1$  generated by  $w_1$ , a distinct branch shall represent  $x_1$  generated by  $w_2$ , and yet another  $x_1$  generated by  $v_1$ : The tree structure does not allow for multiple parents. Branches must then multiply, and more branches be assigned to the same event when produced by distinct 'causes'. The apparent complication is counter-balanced by the logical clarity of the tree structure, which allows analyzing contingencies as in a chess game. Also in chess, the same move may follow distinct moves by the other player, or one's own, but a mental tree representation may facilitate an assessment of the current situation, at the price of some redundancy. When mentally climbing on a branch that corresponds to exactly the same situation of the pieces as another already visited branch, we only need to identify the two and retrieve the configuration of thinner branches, and leaves, already explored.

Inconsistencies may arise if, following Badre & D'Esposito, to the branches of a tree representation, conceived as descriptive of a mental process, one wishes to associate neuronal activity in certain cortical areas so that, for example, value  $x_1$  of variable  $x$  implies specific activity by a particular group of neurons. And one insists on a generically valid correspondence, so that those neurons 'code' for  $x_1$ .

Then either the distinct branches corresponding to  $x_1$  when generated by distinct causes are represented by the activity of distinct groups of neurons, and then the tree hierarchy is clear but the coding is confusing and highly redundant, or they are all represented by the same group, in which case the coding is clear but the hierarchy loses much of its general significance, beyond the individual experimental design, and the simple-minded logical tree of contingencies grows into a mangrove of multi-factorial events, eventually sublimating into a web of interactions of surreal complexity — a cortical network, obviously.

#### 1.4. *Syntactic Trees and Hierarchical Processing in Language*

A domain in which tree representations have been developed to more powerful sophistication and have offered what seems like an essential contribution is, of course, in the description of syntax in natural languages. Such phenomenologically observed syntax is often interpreted, in the various streams of formal linguistics, as the imperfect biological manifestation of an exact underlying structure, which takes the form of an (upside-down) parsing tree. The terminal branches, or leaves, are associated roughly to individual words  $w, x, y, z$  of a natural language, while the non-terminal branches are associated to grammatical constructs  $A, B, C$  that do not appear overtly in natural language, but which are usually construed to have a neuronal representation of some form in our brain. From a start symbol  $S$  one generates a sentence by following not a *single* branch at each node, as in the complexified conditioning paradigms mentioned above, but *multiple* branches, as specified by certain rewrite or production rules. The corpus of sentences, generated by all possible ways of following each production rule, coincides conceptually with the language and can be represented by a gigantic tree, but even a single sentence corresponds to a tree that has as many leaves as, roughly, words — in contrast to single chess games and individual conditioning trials, which do not correspond to flourishing trees, but rather to destitute trees that have lost all leaves but one. There is thus a hierarchy of levels of analysis implicit in each sentence, if described by a parsing tree, which does not necessarily match the hierarchies necessary to parse other sentences, even within the same formal framework.

Chomsky (1955) has famously shown that such frameworks can be classified in a further, abstract hierarchy of frameworks, from unrestricted to context-sensitive to context-free to regular grammars. This was a beautiful achievement, fertile with insightful connections to then-developing computer science. Strictly speaking, it is more of an ordered set of inclusion relations than a *bona fide* hierarchy. (It does not befit a group of oligarchs to be much more numerous than the populace they rule over — if it is unrestricted grammars that are considered to be on top; and it does not belong to the rulers to be simpler-minded and less powerful than their subjects — if vice versa it is regular grammars that are considered to be on top.) In any case, the hierarchy of grammars has not been associated to hierarchical structures in the brain, except perhaps by some fundamentalists who have reputed simpler brains (which they might unwittingly ascribe to non-human primates, or to other mammals) to be capable of cognitive processes equivalent to regular grammars.

Parsing hierarchies, instead, have been associated, in one of the most striking results of applying linguistic theories, with an orderly progression in the severity of the deficits in aphasic patients with different patterns of agrammatism (Friedmann 2001). In particular, among three groups of agrammatic patients, the one most impaired patient could be described as being able to access only the leaves and the thinnest branches of the syntactic tree expressing the structure of a Complementizer Phrase (CP), which, in the usual upside-down scheme, has the CP node at the top and the leaves dropping down at various levels on the left. 13 severely impaired individuals could be described as being able to climb up two more levels, and 5 mildly impaired ones to climb two further levels up, getting almost within sight of the CP node, as it were. These intriguing observations suggest the psychological reality of parsing trees, as ways to order syntactic structures in terms of relative complexity; they do not however, point to a correspondence between those trees and connectional hierarchies in the cortex, even though one is tempted to make that inference, given the long standing neuropsychological association between specific brain lesions and specific behavioral impairments.

The relation between agrammatic impairment and brain lesion is made problematic by the possibility, in general, that the same production rule be represented, in different derivations (in the parsing of different sentences) at different levels of the tree. Then, if different levels of the parsing tree are forced to correspond to different levels in one of the connectional hierarchies of the cortex, either one assumes that a given production rule be potentially expressed by the operation of several different cortical areas, or else one has to assume that the anatomo-functional correspondence is itself variable, from sentence to sentence, with maybe only the initial symbol S, which generates the whole sentence, represented in a stable manner in a mythical spring, somewhere in the frontal lobes, out of which every sentence gushes forth. Neither option is particularly appealing, once made explicit, but in implicit form they may guide our thinking, however unwilling we are to acknowledge it.

## 2. Recursion and Recurrence

Recursion, in whatever form, makes any attempt to impose a rigid hierarchical processing scheme on the cortex even more difficult, in the same plain sense in which social mobility disrupts rigid social hierarchies. Recursion in a weak form arises immediately when one conceives of a natural language as being satisfactorily approximated by the corpus generated by a finite set of production rules, in the sense that the same set of rules is applied at each step  $k$  in the generation of a sentence, however long, and potentially the same terminals (the leaves) can be attached at different steps. This might be dismissed as trivial recursion, but modelers who take seriously the challenge of identifying neuronal mechanisms apt to implement syntactic operators, for example, in terms of fillers and roles (beimGraben *et al.* 2008; Borensztajn *et al.* 2009; Battaglia & Pennartz, in press; see also Namikawa & Hashimoto 2004), devote much of their creativity to dealing with such 'trivia', and appropriately so. This form of recursion can stretch out

across very many steps, especially when syntactic dependences are considered to extend, as they do in real life, beyond individual sentences. It is, to all intents and purposes, infinite recursion.

Recursion takes a stronger form when the set of rules allows for choosing the very same individual rule at different steps, as necessary to model simple complementizer sentences like 'John reports that Mary says that...' or, as in Dante's 13<sup>th</sup> canto of the *Inferno*, 'Cred' io ch'ei credette ch'io credesse che tante voci uscisser, tra quei bronchi...'.

Yet more complicated forms of recursion occur when instead of a linear chain, that, for example, includes two non-terminal elements  $A$  e  $B$  (Verb Phrases and Noun Phrases, say), which take terminal values  $a_1b_1a_2b_2a_3b_3$ , one has an embedded structure like  $a_1a_2a_3b_1b_2b_3$  or even  $a_1a_2a_3b_3b_2b_1$ . Formal grammars that admit these structures are the less restricted ones, but it seems obvious that such structures are needed in order to model natural language without having to resort to byzantine constructs. The design of artificial systems that produce language-like strings endowed with such embedded structures is difficult, and it has been suggested that non-human primates cannot speak because they cannot manage the embedded type of recursion (Fitch & Hauser 2004). Yet there is no convincing evidence that non-human primates can acquire, and speak fluently, natural languages even without embedded structures (Hauser *et al.* 2002). And from the point of view of messing up the correspondence between processing along a fixed anatomical hierarchy and along a dynamically rearranged syntactic tree, embeddings are not needed: The simpler types of non-embedding recursion are sufficient — especially when recursion is expressed along several distinct dimensions. A point about embeddings which is often overlooked is that humans generally have trouble understanding, and rarely produce, embedded structures with more than 3 or 4 levels of embedding. Thus the distinction between infinite recursion in humans and finite in other species, proposed by Hauser *et al.* (2002) does not coincide with that informing the experiments of Fitch & Hauser (2004), between finite levels of embedding in humans and zero in other species.

What the observation of even simple forms of recursion suggests is that one should abandon the hypothesis that, during language production, processing should occur along an ordered hierarchy of cortical areas, whether or not specialized for distinct operations (as 'modules' in the Fodorian sense; Fodor, 1983). Such hypothesis originates in computer science and in the block diagrams of early cognitive psychology, but is completely foreign to the world of cortical information processing, where recurrence is the rule. If humans were to process information along a feed-forward series of stations when they speak, they would be singularly handicapped, given that all other mammals, and other amniotes, and humans when they do not speak, use complex recurrent circuits all the time. It would be very odd if recurrent processing were to be silenced only in order to permit, of all things, a prominently recursive functionality such as language!

It seems, therefore, that excess reliance on artificial intelligence approaches to language, on trying to analyze language as it would be if it had evolved among computers and not among humans, has led astray the search for the neuronal mechanisms of language production, which conceivably have to be found among generic cortical mechanisms. But perhaps with a twist.

### 3. Quantity May Produce Quality

What distinguishes the human cortex from that of other mammals are its dimensions. In the past, it was thought that the human brain was larger than that of other mammals only in relation to body weight because in absolute terms, elephants (5 kg) and whales (8 kg) have larger brains than humans (1.4 kg). Recently, however, Suzana Herculano-Houzel (2009) has discussed the possibility that the human cortex may have more neurons, also in absolute number, than any other mammal (and any other living organism). Her argument is based on the observation that the human cortex appears to scale up linearly with respect to other primates, with an approximately constant density of cells per unit volume; whereas with rodents the scaling is strongly sub-linear: A large rodent like the Capybara has much reduced density, and many fewer cells than expected in a linearly scaled up mouse of that body weight. While the scaling laws for the density of neurons in proboscidea and cetaceans are not known, it is likely, she argues, that they will end up making the largest whale and elephant brains less dense, as with rodents, resulting in total number of neurons in their cerebral cortices around 3 billion, compared to 16 billion in the human cortex. So, the human cortex would be the one with more neurons, after all.

Other quantitative parameters that affect the capabilities of neuronal networks are those that determine their connectivity. The observation of uniform design principles for the cerebral neocortex, across mammalian species and cortical areas, should not be taken to imply that the 'canonical' cortical circuit (Douglas & Martin 1991) is exactly the same, hence tacitly assume it to operate always in the same manner (Rakic 2008). Guy Elston, in fact, has argued now for a number of years that important quantitative differences exist in the number of spines present on the dendrites of pyramidal cells. Focusing on estimating the number of spines on basal dendrites, taken to be indicative of the number of independent recurrent synaptic inputs from other pyramidal cells nearby, he has reported much lower numbers in occipital than in temporal or frontal cortex, e.g., 1,000 vs. 7,000 or 9,000, respectively, in the macaque. Further, the corresponding numbers are all significantly higher in the human cortex, ca. 2,000, 13,000, and 15,000, respectively (Elston *et al.* 2001). Such quantitative differences are normally overlooked in conceptual reasoning, but they can easily produce qualitative differences in the functionality of a neural network.

#### 3.1. Phase Transitions

The physics of phase transitions offers a poignant model for how quantitative differences in some parameter describing a complex system of interacting units can generate major qualitative differences in the collective behavior of the system. This will not be reviewed here, but it suffices to note that conceptual causal reasoning alone, where consequences are logically associated to qualities, with disregard for quantities, would have had difficulties explaining why water is liquid at 33° F, but turns into ice just 2 degrees below, or why a ferromagnetic material can suddenly lose its properties upon heating. Conceptual explanations, sadly still the dominant epistemological paradigm in the cognitive sciences, are



inadequate when dealing with phase transitions. The mathematical techniques originally developed to analyze models intended to describe one particularly complex type of materials exhibiting phase transitions, the so-called spin glasses, have instead been successfully adapted to analyze models of associative memory networks, following the suggestion by John Hopfield (1982). Such networks, like spin glasses, exhibit phase transitions, for example, when their effective ‘temperature’ — a measure of the variability ascribed to noise — changes a bit, but also with changes in their connectivity (Amit *et al.* 1987, Treves & Rolls 1991). Attractor states representing memory items then disappear, if either the effective temperature is too high or the connectivity too low in relation to memory load, and the network enters a phase in which it cannot function as an associative memory. This is the phase transition associated with storage capacity, which we can denote with the critical memory load — the maximum number of memory items  $p_c$  above which associative retrieval fails.

Most studies of formal/mathematical models of associative memory have focused on single networks, which have often been interpreted as representing a patch of cortex, for example,  $1 \text{ mm}^2$ , containing of order  $10^5$  neurons. Valentino Braitenberg, however, has proposed considering the entire cortex, or at least its fronto-temporal associative areas, as an “associative memory machine” (Braitenberg & Schüz 1991), including, in the human brain, of order  $10^5$  patches with  $10^5$  pyramidal cells each (Braitenberg 1978; see Fig. 2). One can consider a mathematical model of such a two-tier associative memory network, in which neurons are grouped into compartments with dense internal connectivity and sparse connectivity between compartments. A model of this type can be called modular, but not in a Fodorian sense, since all the modules, representing real patches but as if they had sharp boundaries, have the same structure and mode of operation. It shows phase transitions, at least the storage capacity phase transition, in that the asymptotic attractor states correlated with each of  $p$  stored memory patterns are only present when both the internal connectivity and the one between modules are sufficiently dense, in relation to  $p$  (O’Kane & Treves 1992).

Specifically, the memory patterns can be defined across the entire network as composed of local patterns stored in each module, which can store  $S$  local attractors if the number of internal connections per unit,  $C_s$ , is sufficient. The number of long range connections per unit,  $C_l$ , has to be sufficient to support the retrieval of the  $p$  stored combination of local attractors, against the interference of all other possible combinations. The minimal values for  $C_s$  and  $C_l$  that allow for successful retrieval, given  $p$  and  $S$ , depend in a complex manner on the parameters characterizing the architecture of the network and the memory representations it encodes (Fulvi Mari & Treves 1998). This makes it convenient to analyze a simplified model, in which the lower tier of the modules with their internal connectivity is replaced by a symbolic representation, in terms of variables which can take multiple values standing for the multiple local attractors of the full model (Fig. 2).

These variables are called Potts units, and can be thought of as tiny vectors pointing towards the multiple directions of a hyper-pyramid in  $S$  dimensions, with its vertex at the origin, a vertex that stands for the inactive state of the local module. Such a Potts version of an associative memory network, which then

explicitly models only the upper tier of the two-tier architecture, has first been analyzed by Kanter (1988) and then by Bollé *et al.* (1991, 1993). They considered discrete Potts units, which can be in a state or another but not partly in a state and partly in another. Graded-response units have been considered later (Treves 2005) and they allow a more realistic modeling of local patch dynamics, including firing rate adaptation, i.e. neural fatigue, a pervasive feature of cortical dynamics.

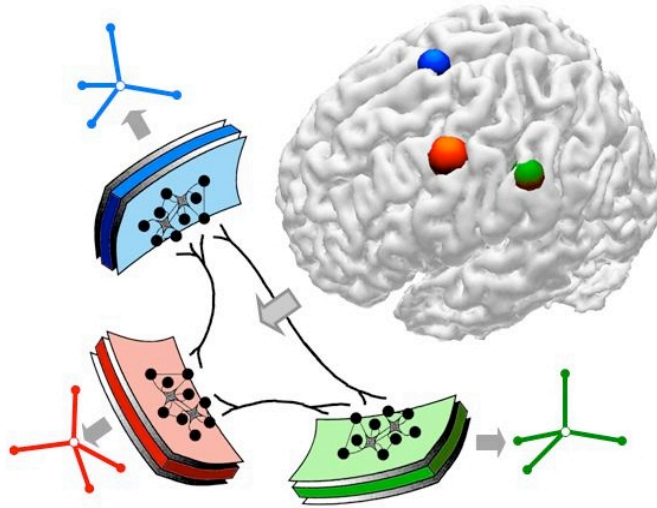


Figure 2: From the cortex to Braitenberg's 'skeleton' model (larger arrow) to the Potts model (small arrows), through abstraction and simplification. Each Potts unit has  $S$  'attractor' states (filled) plus the 'quiescent' state (empty circle).

### 3.2. Latching Dynamics

Including a model of firing rate adaptation leads, in the presence of *correlations* among the global attractors, to a new phenomenon: latching dynamics. Latching dynamics is the hopping of the network from one attractor state to another, where the first, while decaying away due to neuronal fatigue, acts as a cue to retrieve the second, due to their being correlated. The process can be repeated a few times or even indefinitely, in which case one talks of infinite latching.

Latching dynamics are not recursive *per se*. If the transitions from one attractor  $\sigma(n)$  to the next are random, there are no rules of the type  $\sigma(n+1) = \Omega[\sigma(n)]$  being recursively applied at each transition. If the transition probabilities are non-trivially *structured*, however, either sculpted by a learning process or at least embedded in the correlations between attractors, then the transition matrix  $\Omega[\sigma]$  can be regarded as implicitly recursive. It was in fact shown that even for a simple non-structured Potts network the transition probabilities are structured, by the correlations, and 3 distinct classes of transitions can occur between attractor states (Russo *et al.* 2008). More interestingly, it was shown that these transitions cannot be described by a first-order Markov process, as they depend on preceding states, much as words do in natural languages (Russo *et al.* 2011), so one

has to think in terms of long probabilistic dependencies  $\sigma(n+1) = \Omega[\sigma(n), \sigma(n-1), \sigma(n-2), \dots]$ . The Potts network has memory, in its spontaneous gibberish.

### 3.3. *The Phase Transition to Infinite Latching*

We have analyzed latching dynamics in a number of studies (Kropff & Treves 2006, Russo *et al.* 2008, Russo *et al.* 2011) to which we refer for a more detailed description of this behavior of the Potts associative network model, and of its possible relation to neurophysiological observations. The point we want to note here is that latching may never occur, terminate by itself after a few steps, or continue indefinitely. Once the parameters of the Potts network are set, the exact duration of the process, and whether it terminates or not, depend on the exact initial conditions. In a large network, however, the dependence on the initial conditions may become negligible, and if latching terminates after a while it has a well-defined length, dependent only on structural parameters. One can then talk about a phase of finite latching (which may include a region of zero duration, where latching does not even start) and a phase of infinite latching.

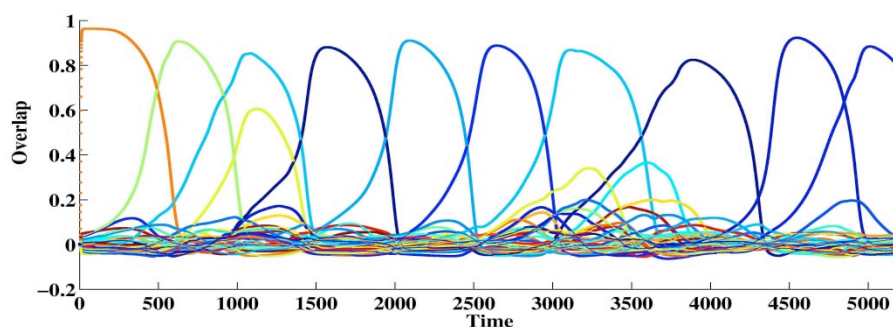


Figure 3: Latching dynamics are shown by plotting the time course of the overlaps of the state of the Potts network with each of  $p$  memory states (in different colors). Ideally, the network hops from attractor to attractor, with rapid transitions and protracted permanence in each attractor. In practice, with many choices of parameters, the observed dynamics are much noisier, a rumble-and-tumble with occasionally several overlaps simultaneously high, and many others significantly above chance level. Here, an input cue is presented at time 0, and then the network is left to its own dynamics. Time steps are in arbitrary units, but interpreting them as msec, or as fractions of msec, may be useful to suggest a correspondence with cognitive time scales.

While our detailed analysis of the boundary between the phases of finite and infinite latching will be discussed elsewhere, it is useful to sketch the scenario that emerges from the possibility of an abrupt phase transition — a distinct phase transition from the storage capacity one, which we had labeled with the critical storage load  $p_c$ . The network in the two phases is identical, and the only difference is in the numerical values of some parameter, for example the number  $S$  of local states of each patch/Potts unit or the connectivity  $C$  of the Potts units (related to the long-range connectivity in the underlying two-tier model,  $C_l$ ). A small change in the value of the parameter then opens the gate for a distinct

emergent property, which is manifest in one phase and not in the other. In the Potts model of the two-tier cortical network there are therefore two critical boundaries. First, there is a boundary between retrieval and non-retrieval phases. The network cannot function as an associative memory because it cannot retrieve a memory item with a partial cue, when the storage load  $p$  is above a value  $p_c$  proportional to  $C$  and to  $S^2$  (Kropff & Treves 2005). Second, there is a boundary between finite and infinite latching. The network latches indefinitely when the memory load is *above* a certain critical value  $p_l$ , because correlations among attractors increase with the memory load. At least in a certain parameter regime, simulations indicate that  $p_l$  does not depend on  $C$  and scales up with  $S$ , approximately linearly. A phase space for the network has to be drawn in (at least) 3 dimensions,  $p$ ,  $C$  and  $S$ , and two orthogonal sections though this phase space are shown in Fig. 4.

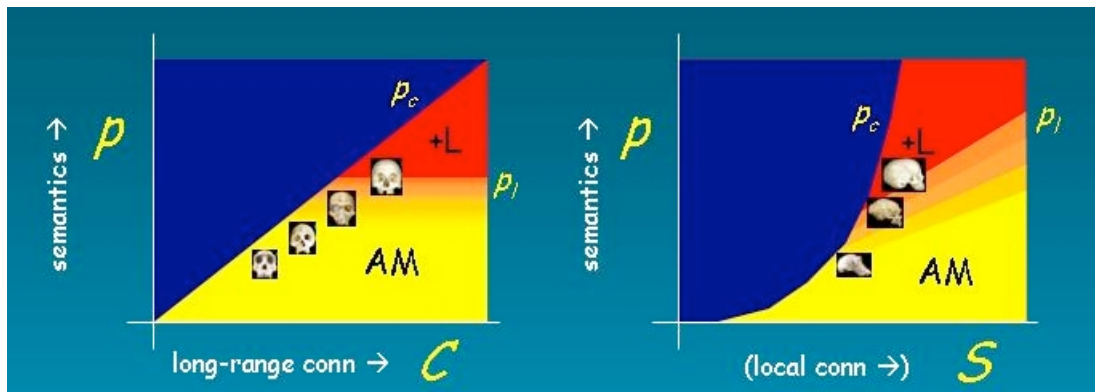


Figure 4: Two schematic cross sections through the 3D phase space of a Potts associative memory. In the  $p$ - $C$  section, left, the network can operate as an associative memory (AM) below a critical value  $p_c$  which is approximately linear in  $C$ . Above a value  $p_l$ , which does not seem to depend on  $C$ , the network also latches indefinitely (+L). A quantitative increase in long-range connectivity, at some point in the evolution of the human brain, may have triggered the emergence of indefinite latching. A similar phase transition (also depicted as a graded change through orange colors, because its exact nature has not been completely determined yet) from the AM to the +L phases may be seen in the  $p$ - $S$  section, right. In this case  $p_c$  is taken to be quadratic in  $S$ , and  $p_l$  linear.

Our on-going analyses indicate that the phase diagram of even simple unstructured Potts networks is not as simple as suggested by Fig. 4. Still, the scenario remains open, that a slowly evolving quantitative increase in the connectivity of the cortex may have suddenly crossed a critical threshold, in the human species, several tens of thousands years ago, that brought the cortical network into a phase characterized by long spontaneous latching sequences, or by their real cortical equivalent, without altering the intrinsic make-up of the network or any of its constituent properties. Latching is an emergent property, or a somewhat more complex set of emergent properties, which emerge when crossing certain thresholds.

#### 4. Conclusion

It is a long shot to extrapolate from transitions produced by randomly generated correlations among attractors, in a crude Potts network, to the recursive concatenation of linguistic structures in speech, even though latching transitions have been shown to display a certain degree of internal complexity (Russo *et al.* 2008). Still, the Potts model indicates the possibility that a recursive mechanism may emerge through a phase transition, in a manner entirely unrelated to the hypothetical appearance, in evolution, of a novel piece of neural circuitry with specific language-adaptable properties (the mythical LAD), or to the refinement of specific connectional hierarchies among cortical areas. The latter may of course encroach on the originally non-hierarchical mechanism and complexify it, but they (the LAD and the hierarchy) may have nothing to do with the emergence of the mechanism. Further studies are needed in order to understand how latching dynamics can be sculpted in a more purposeful manner than by randomly generated correlations, through for example temporally asymmetric synaptic plasticity.

#### References

- Amit, Daniel J., Hanoach Gutfreund & Haim Sompolinsky. 1987. Statistical mechanics of neural networks near saturation. *Annals of Physics* 173, 30–77.
- Amunts, Katrin, Marianne Lenzen, Angela D. Friederici, Axel Schleicher, Patricia Morosan, Nicola Palomero-Gallagher & Karl Zilles. 2010. Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biol* 8, e1000489.
- Amunts, Katrin, Axel Schleicher & Karl Zilles. 2004. Outstanding language competence and cytoarchitecture in Broca's speech region. *Brain and Language* 89, 346–353.
- Badre, David & Mark D'Esposito. 2007. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience* 19, 2082–2099.
- Barbas, Helen. 1986. Pattern in the laminar origin of corticocortical connections. *The Journal of Comparative Neurology* 252, 415–422.
- Battaglia, Francesco P. & Cyriel M. Pennartz. In press. A computational architecture for memory consolidation: A correlational code for episodic memory training hierarchical semantic networks. *Frontiers in Computational Neuroscience*.
- beimGraben, Peter, Dimitris Pinotsis, Douglas Saddy & Roland Potthast. 2008. Language processing with dynamic fields. *Cognitive Neurodynamics* 2, 79–88.
- Bollé, Désiré, Roland Cools, Patrick Dupont & J. Huyghebaert. 1993. Mean-field theory for the Q-state Potts-glass neural network with biased patterns. *Journal of Physics A* 26, 549–562.

- Bollé, Désiré, Patrick Dupont & Jort van Mourik. 1991. Stability properties of Potts neural networks with biased patterns and low loading. *Journal of Physics A* 24, 1065–1081.
- Borensztajn, Gideon, Willem Zuidema & Rens Bod. 2009. The hierarchical prediction network: Towards a neural theory of grammar acquisition. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2974–2979.
- Braitenberg, Valentino. 1978. Cortical architectonics: General and areal. In Mary A.B. Brazier & Hellmuth Petsche (eds.), *Architectonics of the Cerebral Cortex*, 443–465. New York: Raven.
- Braitenberg, Valentino & Almut Schüz. 1991. *Anatomy of the Cortex: Statistics and Geometry*. Berlin: Springer-Verlag.
- Briscoe, Ted. 2000. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language* 76, 245–296.
- Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University/Massachusetts Institute of Technology. [Published in part as *The Logical Structure of Linguistic Theory*, New York: Plenum, 1975.]
- Douglas, Rodney J. & Kevan A.C. Martin. 1991. A functional microcircuit for cat visual cortex. *Journal of Physiology* 440, 735–769.
- Elston, Guy N. 2000. Pyramidal cells of the frontal lobe: All the more spinous to think with. *Journal of Neuroscience* 20, RC95 (1–4).
- Elston, Guy N., Ruth Benavides-Piccione & Javier DeFelipe. 2001. The pyramidal cell in cognition: A comparative study in human and monkey. *Journal of Neuroscience* 21, RC163 (1–5).
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Friedmann, Naama. 2001. Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research* 30, 71–90.
- Fulvi Mari, Carlo & Alessandro Treves. 1998. Modeling neocortical areas with a modular neural network. *Biosystems* 48, 47–55.
- Fuster, Joaquin M. 2009. Cortex and memory: Emergence of a new paradigm. *Journal of Cognitive Neuroscience* 21, 2047–2072.
- Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- 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.
- Herculano-Houzel, Suzana. 2009. The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience* 3, art. 31, 1–11.
- Herculano-Houzel, Suzana, Christine E. Collins, Peiyan Wang, Jon H. Kaas & Roberto Lent. 2008. The basic nonuniformity of the cerebral cortex. *Proceedings of the National Academy of Sciences USA* 105, 12593–12598.
- Hopfield, John J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA* 79, 2554–2558.
- Huggenberger, Stefan. 2008. The size and complexity of dolphin brains — a

- paradox? *Journal of the Marine Biological Association of the UK* 88, 1103–1108.
- Kaas, Jon H. 1997. Topographic maps are fundamental to sensory processing. *Brain Research Bulletin* 44, 107–112.
- Kanter, Ido. 1988. Potts-glass models of neural networks. *Physical Review A* 37, 2739–2742.
- Kropff, Emilio & Alessandro Treves. 2005. The storage capacity of Potts models for semantic memory retrieval. *Journal of Statistical Mechanics: Theory and Experiment*, P08010.
- Kropff, Emilio & Alessandro Treves. 2006. The complexity of latching transitions in large scale cortical networks. *Natural Computing* 6, 169–85.
- Krubitzer, Lea. 1995. The organization of neocortex in mammals: Are species differences really so different? *Trends in Neuroscience* 18, 408–417.
- Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–1706.
- Lorente de Nó, Rafael. 1938. Architectonics and structure of the cerebral cortex. In John F. Fulton (ed.), *Physiology of the Nervous System*, 291–330. New York: Oxford University Press.
- Namikawa, Jun & Takashi Hashimoto. 2004. Dynamics and computation in functional shifts. *Nonlinearity* 17, 1317–1336.
- O’Kane, Dominic & Alessandro Treves. 1992. Short- and long-range connections in autoassociative memory. *Journal of Physics A: Mathematical General A* 25, 5055–5069.
- Rempel-Clower, Nancy L. & Helen Barbas. 2000. The laminar pattern of connections between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function. *Cerebral Cortex* 10, 851–865.
- Rakic, Pasko. 2008. Confusing cortical columns. *Proceedings of the National Academy of Sciences USA* 105, 12099–12100.
- Rockel, A.J., Robert W. Hiorns & Thomas P.S. Powell. 1980. The basic uniformity in structure of the neocortex. *Brain* 103, 221–244.
- Russo, Eleonora, Vijay M.K. Namboodiri, Alessandro Treves & Emilio Kropff. 2008. Free association transitions in models of cortical latching dynamics. *New Journal of Physics* 10, art. 015008.
- Russo, Eleonora, Sahar Pirmoradian & Alessandro Treves. 2011. Associative latching dynamics vs. syntax. In Rubin Wang & Fanji Gu (eds.), *Advances in Cognitive Neurodynamics II*, 111–115. New York: Springer.
- Semendeferi, Katerina, Kate Teffer, Dan P. Buxhoeveden, Min S. Park, Sebastian Bludau, Katrin Amunts, Katie Travis & Joseph Buckwalter. 2010. Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex* 21, 1485–1497.
- Treves, Alessandro. 2003. Computational constraints that may have favoured the lamination of sensory cortex. *Journal of Computational Neuroscience* 14, 271–282.
- Treves, Alessandro. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. *Hippocampus* 14, 539–556.
- Treves, Alessandro. 2005. Frontal latching networks: A possible neural basis for infinite recursion. *Cognitive Neuropsychology* 21, 276–291.

- Treves, Alessandro & Edmund T. Rolls. 1991. What determines the capacity of autoassociative memories in the brain? *Network* 2, 371–397.
- Treves, Alessandro, Ayumu Tashiro, Menno P. Witter & Edvard I. Moser. 2008. What is the mammalian dentate gyrus good for? *Neuroscience* 154, 1155–1172.
- Uylings, Harry B.M., Annelise M. Jacobsen, Karl Zilles & Katrin Amunts. 2006. Left-right asymmetry in volume and number of neurons in adult Broca's area. *Cortex* 42, 652–658.

*Eleonora Russo*  
SISSA  
Sector of Cognitive Neuroscience  
Via Bonomea 265  
34316 Trieste  
Italy  
[russo@sissa.it](mailto:russo@sissa.it)

*Alessandro Treves*  
SISSA  
Sector of Cognitive Neuroscience  
Via Bonomea 265  
34316 Trieste  
Italy  
[ale@sissa.it](mailto:ale@sissa.it)



## “A Running Back” and Forth: A Review of *Recursion and Human Language*

Harry van der Hulst (ed.). 2010. *Recursion and Human Language*. (Studies in Generative Grammar 104) Berlin: De Gruyter Mouton.

by David J. Lobina

An attentive reader of the cognitive science literature would have noticed that the term *recursion* has appeared in myriad publications, and in many guises, in the last 50 or so years. However, it seems to have gained a disproportionate amount of attention ever since Hauser *et al.* (2002) hypothesized (for that is what it was) that this property may be the central and unique feature of the faculty of language.

Indeed, a barrage of publications, conferences, and even critical notes in the popular press about recursion has recently flooded academia. The volume under review here is the result of one such conference — one that was celebrated in 2007 at the Illinois State University — and it offers, or so it says, a compendium of works that tackle this notion from different perspectives.

I will not be following the thematic division underlying this volume as a way to frame the “different perspectives” it advertizes. Rather, this critical note will focus on four distinct senses of the term recursion that can appropriately be applied, or so it will be argued here, to four well-defined theoretical constructs of the cognitive sciences. The formal sciences will, naturally, inform most of this discussion, but the focus of this note will fall on relating the different perspectives of this collection to the four senses of recursion I will outline. Ultimately, this review will press one main point: Contrary to a(n apparently) widespread belief, neatly stated in this book’s back cover, it is simply not true that recursive structures in languages “suggest recursive mechanisms in the grammar” (at least not in the sense that is usually intended in the literature — see *infra*).<sup>1</sup>

The one feature that binds together the four theoretical constructs I will be focusing on is the self-reference property that characterizes recursion — a feature that is quite unrelated to the uses to which this notion can be applied. This self-

---

My gratitude to Noam Chomsky for comments on an earlier version of this paper. This work was partly funded by grants SEJ 2006-11955, 2009 SGR 401, and 2010 FI\_B2 00013.

<sup>1</sup> I will refer to individual chapters by writing the last names of their authors in bold. Thus, the first chapter would be referenced as **Sakel & Stapert**. The editor’s introduction to the volume will be the only exception, as I will refer to it as, plainly, the **Introduction**. When addressing a particular chapter in more detail, the first-mention includes the author’s first names; two chapters did not make it into this review essay for sake of coherence, the one by Hans-Jörg **Tiede** & Lawrence Neff **Stout** and the one by Simon D. **Levy**.



reference property is readily demonstrated by the first connotation I will be considering (the primary meaning in mathematics, in fact), which consists in “defining a function by specifying each of its values in terms of previously defined values” (Cutland 1980: 32); that is, a definition by induction (or recursive definition). The factorial functions ( $n!$ ) offer a standard and rather trivial example:

- (1)            *Def.  $n!$ :*
- a.    if  $n = 1$   $n! = 1$                       *base case*
- b.    if  $n > 1$   $n! = n \times (n-1)!$         *recursive step*

Note that the recursive step involves another invocation of the factorial function. Thus, in order to define the factorial of, say, 4 (i.e.  $4 \times 3!$ ), the function must define the factorial of 3, and so on until it reaches the factorial of 1, the base case, effectively terminating the recursion.

A definition by induction ought not to be confused, however, with a related construct that receives similar denominations, such as an ‘inductive definition’, ‘inductive proof’, or ‘mathematical induction’.

An inductive definition, a mathematical technique employed to prove if a given property applies to an infinite set, proceeds as follows: First, we show that a given statement is true for 1; then, we assume it is true for  $n$ , a fixed number (the inductive hypothesis); lastly, we show that the statement is therefore true for  $n+1$  (the inductive step). If every step is followed correctly, we conclude that the statement is true for all numbers (Epstein & Carnielli 2008). An inductive definition, then, also employs recursion, but it additionally includes an inductive hypothesis. These two constructs should not be conflated, even if they are closely related. In fact, it is important to note that inductive definitions are a central feature of recursive definitions in the sense that the former grounds the latter; that is, the recursive definition of a function is justified insofar as it ranges over the domain established by the inductive definition (Kleene 1952: 260 *et seq.*).<sup>2</sup>

Recursive and inductive definitions are discussed a few times in this collection; this is the case, to a certain extent, for **Introduction**, **Langendoen**, **Kinsella**, and, most notably, in Geoffrey **Pullum** & Barbara C. **Scholz**. The latter take issue with the prevalent belief of what they call the ‘infinite claim’; that is, the claim that, for any language, the set of possible sentences is infinite. Their discussion is framed, I believe, alongside three main themes and it is of interest to have a closer look at them; these are: a critique of the actual ‘standard’ argument supporting infinite claims, the accompanying assumptions, and a number of loosely related *obiter dicta* that I will not discuss in any detail.

The standard argument has three parts, according to them: (i) there are some ‘grammatically-preserving extensibility’ syntactic facts of the kind *I know that I exist, I know that I know that I exist*, etc. (p. 115) that lead us to believe that (ii) there is no upper bound on the maximal length of possible sentences (at least for English); these two facts together, in turn, warrant the conclusion that (iii) the

<sup>2</sup> The mathematical literature contains many different types of recursive functions: the primitive class, the general class, and the partial class, among others. They are all recursively defined functions, but range over different types of objects; whatever objects/relations they subsume should not distract from the fact that recursion remains a central property.

collection of all grammatical expressions in a given language is infinite.

The argument is well-put together as far as it goes, and their main worry falls not on the move from (ii) to (iii) (which is simple mathematics, they tell us), but on the transition from (i) to (ii). Interestingly, they do not tell us what is actually necessary to warrant the troubled transition; instead, they dismiss three different possibilities that could be employed for its justification: the use of an inductive generalization, mathematical induction, or by arguing from generative grammars (the latter they take to be, strictly speaking, systems of rewrite rules only; see *infra*). It is not clear at all that any of these strategies have ever been explicitly employed in the literature in order to support the standard argument — at least not in the sense that **Pullum & Scholz** have in mind. Indeed, the examples they do provide are rather strained; specifically, the connection they make between mathematical induction and a remark by Pinker in the context of a popular science book (see p. 119) seems rather weak.

More interestingly, later on in their paper (p. 124 *et seq.*), they point to the (supposedly widely-held) assumption that languages are collections — in a strong mathematical sense.<sup>3</sup> Given that they take this to be the case, it is no surprise the burden they place upon linguists to prove the infinitude claim. In a similar vein, Terence **Langendoen**'s contribution orbits these very issues, and is ended by urging the field to come to an agreement “upon a basis for determining whether a language is closed under one or more of its iterative size-increasing operations” (p. 145).

Note that the whole issue, then, turns out to revolve around too close a connection between natural languages and mathematical systems, to the point that the infinitude of the former is to be proven by the standards that we impose upon the latter. This is, however, unwarranted. It is certainly true that many linguists have employed mathematical techniques and vocabulary to study natural languages, but these were so used because they were useful — certainly not to reduce linguistic phenomena to abstraction. In fact, the latter play a rather limited role in linguistic explanation, for note that linguists typically focus on informants' grammatical judgements in order to unearth the underlying structure of strings. This kind of study focuses on the structure that a certain mental state — *viz.* the linguistic capacity — imposes upon the strings, and not on the strings themselves in isolation from these judgements.

Ultimately, it seems that these authors confuse the use of mathematical concepts as a useful toolkit for a call to reduce linguistics to mathematics, but no such thing ought to be accepted by the working linguist (I will retake the infinitude claim below).

On another note, one would expect to see the opposite argument (that is, the finiteness of a given language) to be placed under the same burden, but this

---

<sup>3</sup> They rank this assumption as one of four factors that may account for the persistent presence of infinitude claims in the literature, but in fact only provide three (pp. 124–129). The second of these — the connection between recursion and linguistic creativity — rests on an obvious misrepresentation, corrected many times before (see Chomsky 2006: xviv). Roughly, creativity does not rest on the ability to construct new sentences (p. 126); rather, this property points to the fact that linguistic behaviour is generally stimulus-free, and that speakers/hearers have the capacity to understand/produce novel sentences that are appropriate to context.

does not appear to be the case for languages that *prima facie* lack the ‘grammatically-preserving extensibility’ syntactic facts mentioned in (i) (p. 130–131). Surely a similar argument would arise: (i’) There are some syntactic facts of language A that suggests this language lacks grammatically-preserving extensibility structures (such as self-embedding and coordination), which leads to believe that (ii’) there is indeed an upper bound on the maximal length of its sentences; therefore, (iii’) this language is a finite collection of sentences. Clearly, the transition from (i’) to (ii’) is as troubling as that of (i) to (ii) — but only if we grant **Pullum & Scholz’s** (and **Langendoen’s**) burden.

Be that as it may, I now want to argue that none of this has, in actual fact, much to do with the introduction of recursion into linguistics — at least not in the sense that Chomsky has treated this notion.

A second sense of the term recursion has it as a general and central property of algorithms and generative systems. Thus, in the analysis of algorithms discipline, systems of recursive equations have been employed to formalize the notion of an algorithm qua formal object (see, especially, McCarthy 1963) and some scholars have proposed that these recursive equations subsume a specific mapping function, termed a ‘recursor’ (Moschovakis 1998, 2001, Moschovakis & Paschalis 2008). A recursor is said to describe the structure of an algorithm and, in this sense, algorithms *are* recursors. Production systems of rewriting rules also contain recursion as a central property, but not simply in those specific cases in which the same symbol appears on both the left- and right-hand side of a rewrite rule.

Consider, for example, the underlying transformation that converts some structure  $\phi_1 \dots \phi_n$  into some structure  $\phi_{n+1}$ ; the  $\rightarrow$  relation can then be interpreted as “expressing the fact that if our process of recursive specification generates the structures  $\phi_1 \dots \phi_n$ , then it also generates the structure  $\phi_{n+1}$ ” (Chomsky & Miller 1963: 284). This is basically the successor function, one of the primitive class of recursive functions. Whereas the former cases involve an internal application of recursion within production systems, the latter is a global property of collections of rewriting rules qua production systems (see *infra*).

The successor function also underlies what is known as the ‘iterative conception of set’, a process in which sets are “recursively generated at each stage”, a statement that is to be understood as the “repeated application of the successor function”, drawing our attention to the analogy between “the way sets are inductively generated [...] and the way the natural numbers [...] are inductively generated from 0” (Boolos 1971: 223). The current characterization of Merge, the building operation at the heart of the language faculty, as a set-formation operator seems to be akin to this interpretation of recursion (see Chomsky 2008 and Soschen 2008).<sup>4</sup>

This is better understood in the context of the discussion Soare (1996) provides on the state of the art within mathematical logic. Therein, he argues that the field has for long assumed that recursion and computation are synonymous terms (and the same would apply for recursive and computable). This, he argues,

<sup>4</sup> Particularly, it is in this context that Soschen’s (2008: 199) statement regarding how “singleton sets are indispensable for recursion” should be understood.

has resulted in what he calls the Recursion Convention (RC), a state of affairs he has attempted to reverse in subsequent publications. The RC has three parts: (i) Use the terms of the general recursive formalism to describe the results of the subject, even if the proofs are based on the formalism of Turing computability; (ii) use the term Church Thesis to denote various theses; and (iii) name the subject using the language of recursion (e.g., Recursion Function Theory).

Granted, even if it is commonly conceded that a Turing Machine captures the manner in which every conceivable mechanical device computes a calculable function (and it is, furthermore, generally accepted as the best formalization in the field), Turing's model did not in actual fact provide a formalization of what an algorithm *qua* formal object is. Indeed, there is a distinction to be had between formalising an algorithm *qua* a 'model of computation' — that is, an analysis of what actually happens during a computational process — and *qua* an abstract mathematical object. It is to the former construct that Turing's model appropriately applies, while it is the latter that systems of recursive equations and recursors subsume. Furthermore, it is a well-known result that Turing Machines and the partial recursive functions formalism of Church/Kleene (but not the general recursive class) are extensionally equivalent in the sense that both identify the class of computable functions. From the same inputs, both formalisms return the same outputs, albeit in different ways (these 'intensional differences' will be of some importance later on, though). Finally, it is no surprise that the Turing Machine model has been more prominent in cognitive psychology, with its emphasis on real-time processes, while the more abstract characterization of what a computation is — one based on recursion — has found its natural place in theoretical linguistics mainly. Indeed, as Collins (2008) states in the context of an introductory book to Chomsky's thought: "via the Church/Turing thesis, computation is just recursion defined over a finite set of primitive functions" (p. 49).

There is, therefore, a certain consistency in Chomsky's writings if we understand his treatment of recursion in the terms just described. Perhaps rather tellingly, he *has* pointed to the connection between grammatical theory and recursive function theory in many writings (e.g., in Piattelli-Palmarini 1980: 101), which suggests that he may have been influenced by the RC.<sup>5</sup>

Naturally, the whole point of introducing recursion into linguistics was to account for the fact that speakers/hearers show a continuous novelty in linguistic behaviour — a novelty that does not appear to be capped in any meaningful respect. Further, since speakers/hearers cannot possibly store all the possible sentences they understand or utter, the cognitive state accounting for this linguistic behaviour must be underlain by a finite mechanical procedure — an algorithm. This is one of those properties that one would argue are a matter of 'conceptual necessity'. A rather trivial matter, perhaps, but the whole point has been muddled by orbiting issues. Just as **Pullum & Scholz** do, many studies focus on the so-called self-embedded sentences (sentences inside other sentences,

---

<sup>5</sup> This is actually confirmed in personal correspondence with Noam Chomsky (May 2009): "[T]here is a technical definition of 'recursion' in terms of Church's thesis (Turing machines, lambda calculus, Post's theory, Kleene's theory, etc.)", the only one used he's ever used, "a formalization of the notion algorithm/mechanical procedure". Further, he states that he's "always tacitly followed the RC".

such as *I know that I know* etc.) as a way to demonstrate the non-finiteness of language, and given that self-embedding is sometimes used as a synonym for recursive structures (see *infra*), too close a connection is usually drawn between the presence of these syntactic facts and the underlying algorithm of the language faculty. However, even if there were a language that did not exhibit self-embedding but allowed for conjunction, you could run the same sort of argument and the non-finiteness conclusion would still be licensed. These two aspects must be kept separate; one focuses on the sort of expressions that languages manifest (or not), while the other is a point about the algorithm that generates *all* natural language structures.

There is surprisingly little in the collection under review here that touches on these very issues. Both Harry van der Hult's **Introduction** and Arie **Verhagen** do discuss global and local applications of recursion, but not quite in the terms outlined here. **Verhagen** describes two roles for recursion, a specific one that gives rise to long-distance dependencies (supposedly the self-embedded sentences, see below) and a more general one that delineates a mechanism for embedding phrases inside other phrases. As for the **Introduction**, and even though the discussion presented there is framed in terms of rewriting rules, the main points pertain to structures only. Thus, the general application refers to the general embedding of phrases into other phrases, while the specific one refers to those phrases that are embedded within a constituent of the same kind. I will come back to this below, but it is worth pointing out now that this is the closest this collection gets to discussing the central role of recursion in the formalization of an algorithm — a clear shortcoming, given that the RC, systems of recursive equations, recursors, etc form the core of Chomsky's thought on the matter.

There is some tangential discussion regarding Merge by Jan **Koster**. He takes issue with the postulation of a recursive Merge as the syntactic engine that generates linguistic expression. His worries seem to be twofold; on the one hand, a recursive Merge cannot do anything unless in combination "with external, invented cultural objects — lexical items" (p. 289); on the other hand, these lexical items come with specific combinatorial properties that already account for the hierarchical structure that Merge will, then, redundantly generate "once more" (p. 292). The latter is, of course, a valid point that recapitulates the debate between representational and derivational views on theories of grammar. The former is, however, more troubling. Even if we were to grant **Koster** that lexical items *are* cultural inventions, we would do well to remind ourselves that there cannot be any cultural inventions that are not entertained in the mind first, which is perhaps a trivial point. Moreover, we can be sure that external, cultural inventions do not come with "fully-fledged combinatorial properties" (*idem.*) other than the mould that the linguistic capacity imposes — and this is clearly an internalist explanation. More importantly, this bears very little relation to my description of the role of recursion in linguistic theory as a general property of the underlying mechanical procedure.

I have been defending a common thread running through Chomsky's writings, but this is not to say that he has always been consistent — or that the focus has not fallen, on many occasions, on the internal applications of recursion within production systems.

The characterization of this internal application has changed dramatically as the theory progressed. An early characterization of generative grammar divided the computational system into two components: the base (composed of rewriting rules that returned strings with associated phrase markers) and the transformational system (a component that would convert some phrase markers into other phrase markers, preserving structure). In Chomsky (1957), the recursive property of certain rules is ascribed to the latter system, while Chomsky (1965) assigns it to the base component. By the 1970s and 1980s, most of the rewriting rules were in fact eliminated from syntactic theory, perhaps completely so by the time Chomsky (1986) appeared. The latter is an important point, given that most discussions on recursive mechanisms — and this is no exception in this collection — seems to be centered exclusively on rewriting rules, which is rather unfortunate. It should be trivial at this point to remark that recursion as a general property of generative systems remains at the center of linguistic theory regardless of the replacement of production systems with Merge — both, as I have tried to show, are underlain by the successor function.

Nevertheless, it is of interest to discuss recursive rewriting rules to some extent, given the prominence they receive in the literature — and in this collection. Consider the following sample below.

- (2) a. S → NP VP  
 b. NP → D N  
 c. VP → V NP  
 d. NP → N (NP)  
 e. VP → V S

Rules (2d) and (2e) are recursive, as the category on the left-hand side of the arrow is reintroduced on the right-hand side (directly in (2d), indirectly in (2e)). These can generate what I for now will call nested constructions (sometimes called self-embedding), such as noun phrases inside other noun phrases (such as *John's [brother's [teacher's book]] is on the table*), or sentences inside other sentences (as in *John thinks (that) [Michael killed the policeman]*). In general, it is this sort of structures that most linguists have in mind when they talk about recursion.

It is sometimes supposed that nested structures and recursive rules are very closely connected, to the point that nested constructions cannot be generated by anything other than recursive rules. This is mentioned in this collection a couple of times, sometimes with (dubious) references to the mathematical linguistics literature on formal grammars. This is clearly not quite correct, however. To a first approximation, it is worth noting that recursive rules were introduced in order to simplify the grammar. Take a nested string such as  $[a[ab]b]$ , where the  $a$ 's stand for subjects and the  $b$ 's for verbs. This can easily be generated by the employment of rules like (2d) and (2e), while a more complicated generation would involve the repeated application of rules like  $A \rightarrow aB$ ,  $B \rightarrow aC$ ,  $C \rightarrow bD$ ,  $D \rightarrow b$ . It is precisely in this context that Chomsky (1956: 115–116) states that “if a grammar has no recursive steps [...] it will be prohibitely complex”, with danger of reducing it to a list of sentences.

Amy **Perfors** *et al.* (Josh Tenenbaum, Edward Gibson, and Terry Regie)

provide (partial) confirmation for this intuition by employing a qualitative Bayesian analysis to calculate the ideal trade-off between simplicity of grammars (treated as a prior probability) and the degree of fit to a corpus (treated as the likelihood). Even though recursive rules, they tell us, are costly because they predict sentences that are not observed in a corpus (which hurts their goodness of fit; see pp. 161–164), the calculation ultimately returns, perhaps unenlightening, a grammar with recursive and non-recursive rules as the preferred choice. I qualify these results as uninformative because they do not seem to differ from what was being proposed in the 1950s. Granted, this sort of analysis offers a much more formal understanding, but one should not mistake formalization for insight if the issues were already well-understood. Further, there are two aspects of this work that are somewhat troubling. First, the study places too much emphasis on the actual ‘observed’ data found in corpora. These are not to be disregarded, obviously, but linguists ought not to forget that the actual subject matter, that is, the actual phenomenon to be explained, remains the cognitive state that underlies the observed linguistic behaviour (this point about corpora resurfaces in many other contributions). Secondly, it is an obvious point to make that this analysis only applies to those theories that postulate production systems as grammars — those linguists close to the generative framework, though, have long dispensed with them.

Quite clearly, none of this applies to theories that focus on Merge as the central syntactic engine. Moreover, it certainly has very little to do with the general point made *supra*; namely, Chomsky’s leitmotif is based on recursion qua general property of the computational system underlying language, be this a production system or a set-operator like Merge.

Despite it all, it is worth delving into the uses (and abuses) that systems of rewriting rules have been put to as to unearth some (seemingly) widespread mistakes. This will allow me to introduce the third sense I would like to discuss, one that pertains to the study of computational processes, which is the interest of much of applied computer science. At this point, though, it seems “a reasonable conjecture” to claim that at root “there is only one fixed computational procedure that underlies all languages” (Chomsky 1995: 11); a ‘recursive’ Merge in this sense.

It is important to note that there is a significant discontinuity between rewriting rules and linguistic expressions. Technically speaking, rewriting rules only return strings, not structures, which is presumably one of the reasons why rewriting rules were eliminated from linguistic theory (cf. Collins 2008: 58)<sup>6</sup>. It is a point that deserves emphasis, as its neglect hampers clarity. Take Fitch (2010), for instance; therein, he puts forward two problematic claims: Firstly, that a recursive rule has the property of self-embedding (p. 78), and secondly, that it is a “linguistic stipulation” for a self-embedding rule to entail a self-embedded

---

<sup>6</sup> I say “technically” in reference to the historical fact that rewriting rules have always been employed as string substitution operations. It is sometimes stated, however, that a system of rewriting rules *strongly generates* a set of structures, while it *weakly generates* a set of strings, but there is no obvious difference in the actual rules to merit the distinction — apart from the definition. Perhaps this should be rephrased as follows: A computational system such as rewriting rules generates weakly but a system like Merge generates strongly.



structure (p. 80), which I suppose carries over to simply embedding rules and embedded structures.

The first claim is simply not correct. A rule is recursive if there is a self-call, but this is independent of *what* operation is in fact executed. There is a distinction to be had between what a rule does and how it actually proceeds, and it is to the latter than recursion applies. It is this reflexive property that makes the definition of the factorial functions recursive, but there is no sense in stating that there is any embedding whatsoever.

As mentioned, rewriting rules return strings, not structures; *a fortiori*, there is no such thing as a self-embedding rewriting rule. Moreover, Fitch misplaces the long-held stipulation he identifies. In previous models, the rules of the base component would return simple declarative sentences, and these would be converted into more complex structures by the transformational component; the latter were not part of the set of rewriting rules.

The replacement of Merge for production systems involved the postulation of an operation that embeds elements into one another. Merge does this in a bottom-up fashion rather than generating strings in the left-to-right manner of rewriting rules, but both Merge and a production system are recursive devices for the same reason, that is, qua generative systems that are underlain by the successor function.

The conflation, *apropos* recursion, between what an operation does and how it proceeds is rather common in the literature, and this collection of papers is no different. Some contributions (**Karlsson, Verhagen, Kinsella, Harder, Hunyadi**) discuss what they call center-embedding rules, tail-recursive rules, the sort of structures these generate, and their relationship. Much like Fitch, these terms actually refer to the structures themselves, rather than the actual rules. Thus, a center-embedding rule is supposed to generate nested structures in which, say, a sentence is embedded in the middle of a bigger sentence, like in the classic (*The mouse (the cat (the dog bit) chased) ran away*). A tail-recursive rule, on the other hand, embeds elements at the edge of sentences, either on the left-hand side (*John's [brother's [teacher's book]] is on the table*) or on the right-hand side (*The man [that wrote the book [that Pat read in the cafe [that Mary owns]]]*).

These terms, however, have absolutely nothing to do with the recursive character of the rules themselves, only to the type of embedding the resultant expression manifests. A center-embedding rule, after all, is not one in which the reflexive call occurs in the middle of a derivation, but even if it did, this has no substantial consequences. As for tail-recursion, this is a widely-used term in computer science, and it refers to a process in which the recursive call of the algorithm occurs at the very end of the derivation (Abelson *et al.* 1996). Quite clearly, a nested structure on the left-hand side of a sentence cannot be the result of a tail-recursive rule if the derivation process undergoes left-to-right applications of rewriting rules. In a nutshell, these terms refer to specific properties of the structures, not to recursive mechanisms or operations.

Rather surprisingly, some of the aforementioned chapters seem to have a much stronger point in mind. Fred **Karlsson**, following Parker (2006; cited therein), states that 'nested recursion' rules (i.e. center-embedding; **Verhagen**, p. 103 tells us that this is sometimes known as 'true recursion', but no reference is

provided) cannot be reduced to iterations (while tail-recursion can)<sup>7</sup>, a claim that is repeated by Peter **Harder** (p. 239) and, with qualifications, by Vitor **Zimmerer** & Rosemary A. **Varley** (p. 397).

They could not possibly mean this as a general point about computability theory, however. After all, it is a well-established, though often forgotten, result of the formal sciences that all tasks that can be solved recursively can also be solved iteratively (Roberts 2006). Put bluntly, that is, that “all recursive relations can be reduced to recurrence or iterative relations” (Rice 1965: 114). In fact, one of the references mentioned in this collection, albeit indirectly (p. 347), namely Liu & Stoller (1999), offers a framework that provides automatic transformations of any type of recursion into iteration, an “optimization technique” that can cope with the most complex of recursive relations, such as multiple base cases or multiple recursive steps, of which Fibonacci sequences are an example (contrary to what Fitch 2010: 78 seems to think).

Perhaps what these authors have in mind is a much narrower point; namely, the interrelations between recursion and iteration *within* sets of rewriting rules. In this context, James **Rogers** & Marc **Hauser** offer a solid discussion of formal grammars and their potential relevance for the study of behaviour. Still, the formal literature hardly contains a mention of ‘center-embedding recursion’, a term that only seems to appear in some linguistic papers; as I stated above, it tends to appear in the context of recursive rewriting rules, even if in reality it refers to either an embedding operation of a particular kind, or to a certain type of structure.

As for the recursion/iteration equivalence in general terms, let us take the factorial functions we defined recursively above to clarify this point, which brings me to the third sense I would like to focus on. This refers not to the algorithm qua formal object, but to its actual implementation; that is, it is the study of the so-called models of computation. A recursive process, then, is one in which an operation calls itself, creating chains of deferred operations, which is usefully contrasted with an iterative process, wherein an operation reapplies in succession (Abelson *et al.* 1996: 33–34).<sup>8</sup>

The recursive processing (shown on the left-hand side of Table 1) naturally follows from the recursive definition, while the iterative solution (shown on the right-hand side) necessitates a subtle observation. This is simply that factorials can be iteratively computed if we first multiply 1 by 2, then the result by 3, then by 4, until we reach  $n$ . That is, we keep a running product, together with a counter that counts from 1 up to  $n$ . Further, we add the stipulation that  $n!$  is the value of the product when the counter exceeds  $n$ . (NB: The first digit of the iterative solution shows the factorial whose number we are calculating, the second

<sup>7</sup> Further, he incorrectly states, by misunderstanding the discussion in Tomalin (2006: 64), that Bar-Hillel might have reintroduced recursion into linguistics. Rather, Bar-Hillel seems to have been interested in a more precise definitional technique for theoretical constructs. Chomsky (1955: 45) manifests his agreement in spirit, while two years later sees “success along these lines unlikely” (Chomsky 1957: 58).

<sup>8</sup> This is usefully contrasted with the ‘clear’ definitions of recursion and iteration that **Kinsella** offers on page 182. Note that what she actually provides is a clear example of the conflation we discussed *supra*; namely, between embedding (or not) and recursion (otherwise, iteration).

digit is the actual counter and the third is the running product.)

$4 \times (\text{factorial } 3)$	factiter 4 1 1
$4 \times (3 \times (\text{factorial } 2))$	factiter 4 2 1
$4 \times (3 \times (2 \times (\text{factorial } 1)))$	factiter 4 3 2
$4 \times (3 \times (2 \times 1))$	factiter 4 4 6
$4 \times (3 \times 2)$	factiter 4 5 24
$4 \times 6$	

Table 1: Recursive and Iterative Implementations

As the shape of these implementations show, the material kept in memory at any stage differs greatly. In the second line of the recursive processing, the actual operation in course is factorial 2, while what is being kept in memory is  $4 \times (3 \times \dots)$ . This is in great contrast to any stage of the iterative process, as the only things in working memory are the operation in course and the variables it operates upon. Naturally, an iterative process is in general more efficient; still, there exist clear data structures meriting a recursive solution.

Three properties must be met for a recursive solution to be the most natural: (i) the original problem must be decomposable into simpler instances of the same problem; (ii) the sub-problems must be so simple that they can be solved without further subdivision; and (iii) it must be possible to combine the results of solving these sub-problems into a solution to the original problem (Roberts 2006: 8). Of course, recursive structures are naturally (and intuitively) the ideal candidates, but this should not distract from the point just made, namely that there is nothing intrinsically recursive about the factorial class. That is, the suitability of the recursive solution has to do with the nature of the solution itself, and not with the structures themselves. The connection between a structure and a recursive processing is, therefore, an empirical matter to be worked out on an individual basis; it cannot be simply assumed.

There is a great confusion about this in the cognitive sciences. Thus, much of the literature clearly conflates structures and mechanisms, inevitably concluding that recursive structures can only be generated or produced by recursive mechanisms, and *mutatis mutandis* for iterative structures and mechanisms. I have already noted above that a general property of implementations is that any sort of task which can be solved recursively can also be solved iteratively. Indeed, at the most general level, any function or task that can be computed by the partial recursive functions of Church/Kleene (Kleene 1952), that is, a recursor, is computable by a Turing Machine, and the latter *is* an iterator (Moschovakis 1998). Translating this general result into actual processes is no small matter, but the literature provides many cases (see Liu & Stoller 1999, mentioned *supra*).

There is not an awful lot of discussion on real-time recursive processes in the collection under review here. **Perfors et al.** mentioned, in passing, that even though syntax may well be fundamentally recursive (in the sense of the grammar containing recursive rewriting rules), the parser could “usefully employ non-

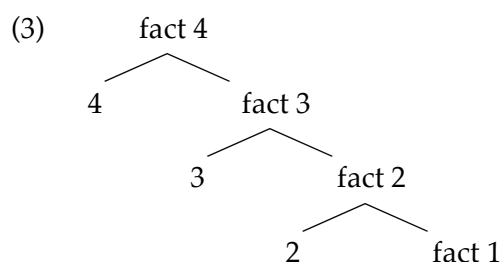
recursive rules" for simpler sentences (p. 170) — a well-taken point about the efficiency of non-recursive rules in processing.

László **Hunyadi** does offer some data regarding possible recursive performance. After correctly stating that recursion and iteration would access (working) memory differently (p. 347), some experimental evidence is provided on the type of prosodic structure associated with some of the structures alluded to earlier (self-embedding and tail-recursion; an 'iterative' structure is also delineated, *viz.* *John is an excellent, cheerful, good-humoured man*). The experiments were rather low-key; subjects were to read some sentences aloud, so that tonal phrases could then be analysed. They found different pitch levels for the embedded sentences in the nested constructions, a phenomenon that was coupled with a 'tonal continuity' — that is, there is a long-distance dependency between two discontinuous segments. For example, for  $[A...[C]...B]$ , the tonal properties of  $[A...B]$  are identical to a continuous  $[AB]$  phrase. Further, this phenomenon is accompanied by three other effects: (i) there is no lowering of the pitch contours in  $C$  (a general tendency called 'downdrift'), (ii) the phrase  $C$  is realized in a different pitch, and (iii) there is an 'upstep' of  $B$  to the initial pitch of  $A$ .

**Hunyadi** sees this process as a clear example of recursion, given that the tonal properties of  $A$  must be kept in memory during  $C$ , so that they can be restored at  $B$ . This, **Hunyadi** believes, is a direct probe of memory — a 'bookmark effect'. Further analyses with tail-recursive and iterative structures show that the bookmark effect does not appear, which would suggest a principled distinction between these and self-embedded structures.

Despite couching the whole discussion in terms of the computational principles of recursion, tail-recursion and iteration, these results appropriately describe structural properties of prosody (or grouping in general; see pp. 361–365), but not its actual production (let alone the underlying grammar). That is, the different memory loads that recursive and iterative *processes* would incur was, in actual fact, not probed at all — there is a distinction, after all, between probing structures and probing mechanisms. We can doubtless be certain that the prosodic structure is, roughly, isomorphic to the syntactic structure, but not much follows about the underlying processing mechanisms.

Furthermore, this contribution introduces some confusion regarding the relationship between recursion and hierarchy (and iteration), and it might be worth our time to clarify it. Take computer science, a discipline which also employs 'trees' in order to represent nonlinear data structures. Computer scientist Donald Knuth certainly echoes a widely-held view when he writes that "any hierarchical classification scheme leads to a tree structure" (Knuth 1997: 312); more importantly, we need to understand his contention that "recursion is an innate characteristic of tree structures" (*idem.*, p. 308). By 'innate', here, is probably meant intrinsic, and we can clarify what this means by providing a graphic representation of the recursive implementation of the factorials, as shown here:



Note that the hierarchical structure directly stems from the fact that the implementation is underlain by a two-equation system: A variable plus a self-call, and it is the latter that expands into the base case, effectively terminating the recursion and the overall computation. It is this specific characteristic that explains why this type of hierarchy, a binary tree, automatically results from a recursive implementation.

This hierarchy, however, is among the operations, and not the data structures. There is no sense in stating, by looking at the tree, that the factorial of 3 is embedded into the factorial of 4. This would amount to a definition of a structure in terms of how it is generated, but why do that? After all, most people are taught at school that the factorial of 4 is calculated by multiplying 4 by 3, then by 2, and finally by 1, and this magically eliminates the once-perceived embedding.

There is certainly a difference between representing a hierarchy of operations and representing a complex object; the factorials example is meant to illustrate that a recursive implementation automatically results in a binary hierarchy, but that one cannot necessarily infer the former from the latter.

It is also worth pointing out that an implementation is a real-time computational process, and we are therefore on a different level of analysis than when discussing, say, Merge. Granted, linguistic expressions also exhibit a binary tree structure, and it is certainly the case that Merge effects this geometry, but crucially it does not do so in the way of a recursive implementation. Recursive generation (successor function) and recursive implementations are different things, even if they may result in similar 'forms'.

It is to the structural 'forms' the language faculty generates that we move onto now, the fourth and last sense of recursion. Recursive data structures are defined by the U.S. National Institute of Standards and Technology as any object or class 'that is partially composed of smaller or simpler instances of the same data structure'; that is, any structure which includes an abstraction of itself (an X within an X). The prototypical cases here are the 'trees within trees' so familiar to generative grammar. It is important to establish what the X in the 'X within an X' is, so as to identify a recursive structure that is in fact of some relevance.

Note that this is a definition that focuses on properties of structures only, independently of the operations/mechanisms that generate/process them. There is nothing odd about this, Chomsky & Miller (1963) defined certain constructions in these very terms; they defined the tail-recursive sentences as either right- or left-recursive (depending on the direction of the embedding), and offered the term self-embedding (still used today) for what some call the center-embedding constructions (p. 290).

It is to actual structures and their properties that a significant amount of

papers in this collection focus on. One set of papers focuses on languages that, *prima facie*, lack self-embedding sentences; I will focus on these first.

Jeanette **Sakel** & Eugenie **Stapert**, then, review the data presented by Daniel Everett on Pirahã, an Amazonian language claimed to lack any type of self-embedding. There are two main points here; one is that Pirahã lacks ‘mental state’ verbs (verbs like *think*, *believe*, etc); *a fortiori*, there is no outright clausal embedding, but simple juxtaposition of individual sentences. The correctness of the latter rests on the status of the verbal suffix *-sai*, a marker that Everett, in earlier work he now considers mistaken, classified either as a nominalizer or as a clausal embedding indicator (see p. 5). Ultimately, **Sakel** & **Stapert** support Everett’s contemporary analysis of this suffix as a single marker of semantic cohesion between parts of discourse. Sauerland (2010), however, offers some experimental data that might cast some doubt on this. After carrying out a maximum pitch analysis on the two conditions the *-sai* marker would appear, according to Everett’s earlier study, Sauerland found that the pitch level in the nominalizer condition was indeed much greater than in the clausal condition, indicating that there *are* two versions of this marker, one of which marks embedding.

Sauerland’s methodology is an interesting one, and can complement more traditional ways of determining whether a language exhibits self-embedding. Marianne **Mithun** lists some of the usual formal features that languages with self-embedding manifest (*viz.* complementizers, omission of co-referential arguments, non-finite verb forms; p. 23) and provides an analysis of a variety of languages, such as central Alaskan Yup’ik, Mongolian Khalkha, and North American Mohawk. Apparently, these three languages exhibit some kind of self-embedding, but in different ways, which suggests, to **Mithun** at least, that this feature manifests cross-linguistic gradience (*sic.*) and variation (p. 39). Perhaps this variation is present within individual languages too. Ljiljana **Progovac** seems to suggest this much for English, given the impossibility of nesting root small clauses (e.g., *Me first, Case closed*; p. 193) into other root small clauses. It is to be supposed that even if this is correct, it is so for a small part of the grammar, leaving other claims (*viz.* the presence of self-embedding else-where) virtually uncontested.

**Karlsson**, on the other hand, offers a typology of recursive and iterative structures (two and six types, respectively; see pp. 43–49 for definitions and examples) based on a quantitative analysis of spoken corpora. By recursive structures he means self-embedding and tail-recursion, and the central claim of his paper is that this sort of corpora analysis provides qualitative data — in the form of ‘constraints’ — that explain why recursive structures are so rare in spoken language. **Karlsson** then concludes that multiple nesting is an artificial feature of language that “arose with the advent of written language” (p. 64) — it is not a central feature of language.

A similar rationale informs Ritva **Laury** & Tsuyoshi **Ono**. Therein, they provide a corpora analysis of conversations conducted in Finnish and Japanese, reaching similar results (and conclusion): Nested constructions are not very common in spoken Finnish and Japanese (pp. 84–55); therefore, recursive structures cannot be a central property of language (I will come back to this presently).

Another set of papers, on the other hand, focus on self-embedding outside

syntax. Thus, Eva **Juarros-Daussà** argues that there is a restriction (what she calls the two-argument restriction) that prevents argument structure (i.e. the predicate with its lexically encoded arguments) to be truly recursive. Quite clearly, however, she is not arguing against the possibility that an element may well be embedded inside an element of the same type (which automatically makes a structure recursive). Rather, she is suggesting that this embedding cannot go on into infinitude (a slightly different matter) — that is, she is arguing for the finitude of argument structure (p. 253). Similarly, Yury A. **Lander** & Alexander B. **Letuchiy** provide data from a Northwest Caucasian language, Adyghe, that seems to allow self-embedding within its verb forms.

On a much grander scale, Harry **van der Hulst** discusses self-embedding in phonology, a topic that has generated some heated debate (as he discusses). Phonological structure is clearly hierarchical, but whether it also manifests self-embedding is rather controversial. This chapter defends the controversial view (phonology is recursive), and the overall idea seems to be that given that recursive structures are principally semantic phenomena (a manner of organizing information), there must be an isomorphic structure in morphology (p. 303). The remainder of the chapter offers a long discussion of phono-morphotactic structure, phonotactic structure, and prosodic structure, concluding that there is, after all, self-embedding at the syllable/foot, word, phrase and prosodic level. Pretty grand claims, and it will certainly be interesting to see what the literature makes of it (a thorough discussion of these issues is beyond the present review).

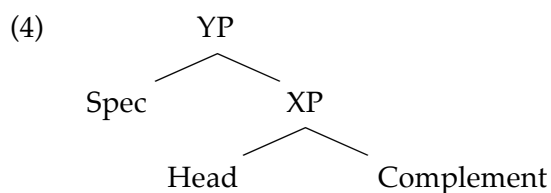
It will have been noticed that I have discussed all these papers in the context of structures only. Indeed, the study of self-embedded structures in natural language is an important one, but it ought to be clear that this phenomenon tells us much more about semantics than about syntax. Such structures, it is clear, provide the linguistic system with a way of “organizing and constraining semantic information”, and their distribution appears to be construction- and language-specific (Hinzen 2008: 358–359).

Once the dubious connection between these structures and specific rewriting rules is disregarded, it is not at all clear why some contributors believe that self-embedding cannot be converted into other types of phrases — a claim that is in fact explicitly denied in **Kinsella** (p. 188). Therein, Anna **Kinsella** makes clear that languages like Pirahã, even if they really do not manifest self-embedding, do not come at an ‘expressive’ loss to their speakers. That is, there is no reason to believe that Pirahã cannot “express [similar] concepts using alternative means”. Indeed, a self-embedded sentence such as *The mouse [the cat [the dog chased] bit] ran away* seems to be easily converted into either *The dog chased the cat that bit the mouse that ran away* (which some would call, I suppose, tail-recursive) or *The dog chased the cat and the cat bit the mouse and the mouse ran away* (a type of iterative structure, according to **Karlsson**).

Furthermore, there is a lot of interesting work regarding the concomitant properties that self-embedded structures exhibit, which range from their role in language acquisition and their cross-linguistic distribution, to their connection to the conceptual system (see, for example, Roeper 2009). Be that as it may, Merge remains a simple, recursive generator for reasons that lie elsewhere — the presence (or not) of self-embedded structures in a particular language is an ancil-

lary matter.

As a final point in this lengthy review, I might as well mention that there are, in fact, grounds to believe that language manifests a much more general type of recursive structure. At the appropriate level of abstraction, a structure that contains an instance of itself (i.e. an X within an X) appears to be a feature of any type of syntactic structure. That is, every syntactic phrase, as Moro (2008: 68) shows, accords to the same geometry, an asymmetric structure [Specifier [Head – Complement]]. This is shown below:



Therefore, a Complementizer Phrase (*viz.* the top node of a clause) is a complex [S[H–C]] structure composed of a number of architecturally-equivalent but simpler [S[H–C]] structures. As Moro (2008: 205 *et seq.*) shows, all human languages appear to follow this scheme, despite some variation in the linear order. Linear order is not the key property; rather, the central point is the basic hierarchical configuration: S is always more prominent than [H–C] and H is always more prominent than C.<sup>9</sup>

At this level, then, structural recursion appears to come for free, but remains an interesting and surprising fact about language. It in fact identifies natural language as a subcategory of infinite systems, one that manifests a specific type of embedding: endocentric and asymmetric X structures. As such, category recursion is a subtype of structural recursion (in the same way that self-embedding is a subtype of general embedding), and it is perhaps in this sense that contemporary debates on the universality of embedding ought to be understood.

Certainly, an extensive terminological clean-up is in order, as much of the nomenclature currently in use (such as ‘true, nested or center-embedding recursion’, ‘tail-recursion’, ‘self-embedding rules’, *et alia*) is likely to create confusion rather than anything else.

### Epilogue

It appears that the word recursion entered the English language in the 17<sup>th</sup> century as an adaptation of the past participle of the Latin verb *recurrere* ‘to go back’. Thus in his *An English Expositor* (1616) — that compendium of the “hardest words” — John Bullokar defined *recursion* as “a running back”. A convoluted term it remains, but for different reasons.

<sup>9</sup> The S–H–C schema invokes X-bar configurations. However, current linguistic theory doubts the existence of the specifier position. If so, the overall architecture would be something like this: [... Head ... (Compl) ... [... Head ... (Compl) ...] ...]. The point I am making still applies; that is, this sort of general recursive structure is present in all languages, independently of the most usual form of self-embedding.



This critical note has attempted to outline four contemporary senses of this term that appropriately applies to well-established theoretical constructs of the cognitive sciences. An attempt was made to encompass the material of this collection around these four connotations as to elucidate a number of issues. Naturally, many topics went untreated, and the focus of this review has befallen on two main points.

Firstly, I have claimed that Chomsky, like many in the mathematical literature, takes recursion to be a central property of what a mechanical procedure is. Despite the different applications recursion has received within his vast output, a recent paper states that linguistic “competence is expressed by a generative grammar that *recursively enumerates* structural descriptions of sentences” (Chomsky 2006: 165; my emphasis) — a very close statement to the spirit of the Recursion Convention.

Secondly, I have tried to show that there is a clear conflation between, on the one hand, recursion and (self)-embedding and, on the other, recursive structures and recursive mechanisms. All these should in fact be kept separate unless there are principled reasons (and there might well be) to link them. Their connection, however, cannot be simply assumed.

It is rather clear that the present collection completely disregards the first point, while being guilty, for the most part, of the second. Perhaps we can concoct an explanation for why this collection so utterly fails to address Chomsky’s actual introduction of recursion in linguistics — the overarching effect of one paper: Hauser *et al.* (2002).

It is undeniable that this paper has generated an incredible amount of discussion, but recursion was certainly not its main topic; indeed, to a certain extent, it received a rather indefinite characterization. This has had the unfortunate result that many recent publications on the role of recursion in cognition (and this is true for many of the contributions of the collection under review) come up with rather outlandish definitions, which are then loosely related to the aforementioned piece, even if on closer inspection, the actual work presented has very little to do with it — or more importantly, with Chomsky’s leitmotif.<sup>10</sup>

Unfortunately, the literature is steadily moving towards an increasingly confused study of recursive structures in conflation with mechanisms, obscuring what ought to be a rather straightforward and uncontroversial point: The centrality of recursion within the formalization of the mechanical procedure that underlies the language faculty.

## References

Abelson, Harold & Gerald J. Sussman with Julie Sussman. 1996. *Structure and Interpretation of Computer Programs*. Cambridge, MA: MIT Press.

---

<sup>10</sup> Moreover, there are good reasons to believe that both Hauser and Fitch have a different view (and, in my opinion, an incorrect one) to Chomsky on what recursion actually is; namely, a self-embedding operation (see Hauser 2009 and Fitch 2010).

- Boolos, George. 1971. The iterative conception of set. *The Journal of Philosophy* 68, 215–231.
- Chomsky, Noam. 1955. Logical syntax and semantics: Their linguistic relevance. *Language* 31, 36–45.
- Chomsky, Noam. 1956. Three models for the description of language. In *IRE Transactions of Information Theory IT-2*, 113–124.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.
- Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2006. *Language and Mind*, 3<sup>rd</sup> edn. Cambridge: Cambridge University Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.
- Chomsky, Noam & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In Robert D. Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, 269–322. New York: Wiley.
- Collins, John. 2008. *Chomsky: A Guide for the Perplexed*. London: Continuum.
- Cutland, Nigel. 1980. *Computability: An Introduction to Recursion Function Theory*. Cambridge: Cambridge University Press.
- Epstein, Richard & Walter Carnielli. 2008. *Computability: Computable Functions, Logic, and the Foundations of Mathematics*. Socorro, New Mexico: Advanced Reasoning Forum.
- Fitch, W. Tecumseh. 2010. Three meanings of recursion: Key distinctions for biolinguistics. In Richard Larson, Viviana Déprez & Hiroko Yamakido (eds.), *The Evolution of Human Language*, 73–90. Cambridge: Cambridge University Press.
- Hauser, Marc D. 2009. Origin of the mind. *Scientific American* 301, 44–51.
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The language faculty: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hinzen, Wolfram. 2008. Prospects for an explanatory theory of semantics. *Biolinguistics* 2, 348–363.
- Kleene, Stephen C. 1952. *Introduction to Metamathematics*. Amsterdam: North Holland Publishing Co.
- Knuth, Donald. 1997. *The Art of Computer Programming* (3 vols.). Upper Saddle River, NJ: Addison-Wesley.
- Liu, Yanhong A. & Scott D. Stoller. 1999. From recursion and iteration: What are the optimizations? *SIGPLAN Notices* 34, 73–82.
- McCarthy, John. 1963. A basis for a mathematical theory of computation. In Paul Braffort & David Hirshberg (eds.), *Computer Programming and Formal Systems*, 33–70. Amsterdam: North-Holland Publishing Co.

- Moro, Andrea. 2008. *The Boundaries of Babel*. Cambridge, MA: MIT Press.
- Moschovakis, Yiannis N. 1998. On founding the theory of algorithms. In Harold G. Dales & Gianluigi Oliveri (eds.), *Truth in Mathematics*, 71–104. Oxford: Clarendon Press.
- Moschovakis, Yiannis N. 2001. What is an algorithm? In Björn Engquist & Wilfried Schmid (eds.), *Mathematics Unlimited: 2001 and Beyond*, 919–936. Berlin: Springer.
- Moschovakis, Yiannis N. & Vasilis Paschalis. 2008. Elementary algorithms and their implementations. In S. Barry Cooper, Benedikt Löwe & Andrea Sorbi (eds.), *New Computational Paradigms*, 81–118. Berlin: Springer.
- Piattelli-Palmarini, Massimo. 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge and Kegan Paul.
- Rice, Gordon. 1965. Recursion and iteration. *Communications of the ACM* 8, 114–115.
- Roberts, Eric. 2006. *Thinking Recursively with Java*. Hoboken, NJ: Wiley.
- Roeper, Tom. 2009. *Microscopic Minimalism*. Boston University Plenary Address.
- Sauerland, Uli. 2010. Experimental evidence for complex syntax in Pirahã. Ms., Zentrum für allgemeine Sprachwissenschaft, Typologie und Universalienforschung, Berlin. [<http://ling.auf.net/lingBuzz/001095>]
- Soare, Robert. 1996. Computability and recursion. *The Bulletin of Symbolic Logic*, 2, 284–321.
- Soschen, Alona. 2008. On the nature of syntax. *Biolinguistics*, 2, 196–224.
- Tomalin, Marcus. 2006. *Linguistics and the Formal Sciences*. Cambridge: Cambridge University Press.

David J. Lobina  
Universitat Rovira i Virgili  
Departament de Psicologia  
Centre de Recerca en Avaluació i Mesura de la Conducta  
Ctra. de Valls s/n  
43007 Tarragona  
Spain  
[davidjames.lobina@urv.cat](mailto:davidjames.lobina@urv.cat)

# *Biolinguistics*

Volume 5, Issue 1–2

Winter/Spring 2011

## TABLE OF CONTENTS

### ★ EDITORIAL ★

- 001 Biolinguistic Perspectives on Recursion:  
Introduction to the Special Issue  
Uli Sauerland  
*ZAS, Berlin*  
Andreas Trotzke  
*Universität Konstanz*
- 010 Learning Recursion: Multiple Nested and  
Crossed Dependencies  
Meinou H. de Vries  
*Vrije Universiteit, Amsterdam*  
Morten H. Christiansen  
*Cornell University*  
Karl Magnus Petersson  
*MPI for Psycholinguistics, Nijmegen*
- 036 What in the World Makes Recursion so  
Easy to Learn?  
Fenna H. Poletiek  
*Leiden University*
- 043 Recursion in Language: A Layered-  
Derivation Approach  
Jan-Wouter Zwart  
*Rijksuniversiteit Groningen*
- 057 The Acquisition of Recursion: How Formalism  
Articulates the Child's Path  
Tom Roeper  
*University of Massachusetts, Amherst*
- 087 The Neural Basis of Recursion and Complex  
Syntactic Hierarchy  
Angela D. Friederici  
*MPI for Human Cognitive and Brain  
Sciences, Leipzig*  
Jörg Bahlmann  
*University of California, Berkeley*  
Roland Friedrich  
*Humboldt Universität, Berlin*  
Michiru Makuuchi  
*MPI, Leipzig*
- 105 Implicit Artificial Syntax Processing: Genes,  
Preference, and Bounded Recursion  
Vasiliki Folia  
*MPI für Psycholinguistik, Nijmegen*  
Christian Forkstam  
*Universidade do Algarve, Faro*  
Martin Ingvar  
*Karolinska Institutet, Stockholm*  
Peter Hagoort  
*MPI für Psycholinguistik, Nijmegen*  
Karl Magnus Petersson  
*MPI for Psycholinguistics, Nijmegen*
- 133 An Uncouth Approach to Language  
Recursivity  
Eleonora Russo & Alessandro Treves  
*SISSA, Trieste*

### ★ REVIEWS ★

- 151 "A Running Back" and Forth: A Review of  
*Recursion and Human Language*  
David J. Lobina  
*Universitat Rovira i Virgili, Tarragona*



biolinguistics

ISSN 1450-3417

*Biolinguistics* 5.1–2: 001–169, 2011  
<http://www.biolinguistics.eu>