Dendrophobia in bonobo comprehension of spoken English

Citation for published version:
Truswell, R 2017, 'Dendrophobia in bonobo comprehension of spoken English', Mind and Language, vol. 32, no. 4, pp. 395-415. https://doi.org/10.17617/2.2248195, https://doi.org/10.1111/mila.12150

Digital Object Identifier (DOI):
10.17617/2.2248195
10.1111/mila.12150

Published In:
Mind and Language

Dendrophobia in bonobo comprehension of spoken English*

Robert Truswell

Abstract

Comparative data from Savage-Rumbaugh et al. (1993) concerning the comprehension of spoken English requests by a bonobo (Kanzi) and a human infant (Alia) is consistent with Fitch's (2014) hypothesis that humans exhibit *dendrophilia*, or a propensity to infer and manipulate hierarchical tree structures to a greater extent than other species. This body of data avoids many pitfalls in interpreting results of relevant Artificial Grammar Learning experiments, and therefore complements those experiments. However, findings from language acquisition suggest that the term *dendrophilia* is misleading, in that human infants do not show an initial preference for certain hierarchical syntactic structures. Infants are slow to acquire and generalize the hierarchical structures in question, but they can eventually do so. Kanzi, in contrast, is *dendrophobic*: even though his nonhierarchial strategy impairs comprehension, he never acquires the hierarchical structure.

*Thanks to Jim Hurford, Ray Jackendoff, Tim O'Donnell, Sarah O'Neill, Ian Roberts, Kenny Smith, Maggie Tallerman, Charles Yang, and Moira Yip, and to audiences at Tufts, Groningen, York, Edinburgh, and Essex, for comments on earlier versions of this material.
Address for correspondence: Linguistics and English Language, The University of Edinburgh, Dugald Stewart Building, 3 Charles St, Edinburgh, EH8 9AD, UK

Email: rob.truswell@ed.ac.uk

1. Introduction

Much of our knowledge is hierarchically organized. A major type of hierarchally organized information concerns the relation between complex objects and their component parts, or **constituents**, represented diagrammatically using **tree diagrams**. The centrality of such structured knowledge representations to many aspects of human cognition has been recognized since at least Lashley (1951). Models of human language (Chomsky 1957), planning (Miller et al. 1960), vision (Marr 1982), music (Lerdahl & Jackendoff 1983), and several other domains routinely refer to such structures. This led Fitch (2014) to formulate his **Dendrophilia hypothesis**, as a general claim about human processing of sequential strings:

> 'Humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species.' (Fitch 2014:352)

Fitch's hypothesis is stated as a matter of degree: humans do not necessarily see trees everywhere, and nonhuman animals may relate strings to hierarchical structures in some cases, but humans are more ready to infer tree structures from strings. Conversely, we may hypothesize that nonhuman animals are **dendrophobic** to the extent that they fail to infer optimal, hierarchically structured representations of sets of strings.

In this respect, the Dendrophilia hypothesis differs from similar claims made in Hauser, Chomsky & Fitch (2002). Hauser et al. advanced the 'recursion-only' hypothesis, that FLN (the 'faculty of language in the narrow sense', or the set of species-specific and domain-specific components of the language faculty) contains only the capacity to recursively embed constituents within larger constituents. Whereas the recursion-only hypothesis is absolute and domain-specific (recursive computation is unique to humans and unique to language), the Dendrophilia hypothesis is relative and domain-general (humans are more likely to use hierarchical structural representations, across multiple domains).

Both the recursion-only hypothesis and the Dendrophilia hypothesis imply a series of comparisons, across species and domains. Hauser et al. already consider recursive aspects of numerical cognition in humans and other species (pp.1576–7), and in their conclusion (p.1578) they list a number of other domains where cross-species comparison may well contribute usefully to evaluation and refinement of the recursion-only hypothesis. Pinker & Jackendoff (2005), and Jackendoff & Pinker (2005), pursue this further with a simple demonstration of apparent recursive constituent structure in human visual cognition. Although there is still ongoing debate over this question, such arguments, as well as the work on planning, music, and vision listed above, strongly suggest that hierarchical constituent structure is not unique to language, contrary to the recursion-only hypothesis. Indeed, Hauser et al. contemplate this possibility in their conclusion:

> If … one entertains the hypothesis that recursion evolved to solve other computational problems such as navigation, number quantification, or social relationships, then it is possible that other animals have such abilities… If we find evidence for recursion in animals, but in a noncommunicative domain, then we are more likely to pinpoint the mechanisms underlying this ability and the selective pressures that led to it. This discovery, in turn, would open the door to another suite of puzzles: Why did humans, but no other animal, take the power of recursion to create an open-ended and limitless system of communication? Why does our system of recursion operate over a broader range of elements or inputs (e.g., numbers, words) than other animals? One possibility, consistent with current thinking in the cognitive sciences, is that recursion in animals represents a modular system designed for a particular function (e.g., navigation) and impenetrable with respect to other systems. During evolution, the modular

> and highly domain-specific system of recursion may have become penetrable and domain-general. This opened the way for humans, perhaps uniquely, to apply the power of recursion to other problems. (Hauser et al. 2002: 1578)

The Dendrophilia hypothesis can be seen as embodying this train of thought. If tenable, it is still strong enough to contribute to our understanding of the uniqueness of human language, the question motivating Hauser et al. (2002): human language is apparently unique among communication systems in allowing unbounded, novel pairing of complex forms recursively constructed from simpler elements, with complex meanings recursively constructed from the meanings of the simpler elements. In this way, an ability to manipulate recursively generated hierarchical constituent structures underpins a distinctive feature of human language.

This paper is a contribution to the assessment of the Dendrophilia hypothesis with respect to the core domain of **grammar induction**, or inference of rule systems underpinning strings of sounds. A substantial comparative literature on artificial grammar learning has emerged, following Fitch & Hauser's (2004) demonstration that humans can learn to recognize sequences corresponding to the rule schemata $(ab)^n$ and $a^n b^n$, while cotton-top tamarins can only learn the former. Fitch (2014) discusses this result as support for the Dendrophilia hypothesis. However, its pertinence to a literal interpretation of Dendrophilia is unclear, for reasons discussed in Section 2.

In large part, this unclarity reflects limitations of the experimental paradigm as applied to nonhuman subjects. This paper offers a complementary source of evidence supporting Fitch & Hauser's conclusion, by examining the behavioural responses of a nonhuman primate to spoken English, a hierarchically structured, semantically interpreted natural language. We can infer aspects of the subject's interpretation of an utterance from his behaviour, and aspects of the grammatical representation of the utterance from that interpretation.

The data come from a corpus of 660 utterances directed to the bonobo Kanzi, together with descriptions of Kanzi's behaviour in response (Savage-Rumbaugh et al. 1993). We also discuss Savage-Rumbaugh et al.'s parallel corpus of requests directed at Alia, a human infant, over a 6-month period when she was 18–24 months old. Kanzi's generally impressive performance dips significantly when he is asked to act on multiple objects described by a coordinate noun phrase. Alia showed no such problem. I argue that this is an instance of dendrophobia on the part of Kanzi, that is, failure to infer a hierarchical grammatical structure even when correct interpretation depends on that structure.

The paper is structured as follows. Section 2 reviews recent artificial grammar learning experiments. Section 3 discusses Kanzi's comprehension of spoken English and relates it to those experiments, showing how the two approaches reinforce each other. I show that Kanzi fails to interpret coordinated noun phrase objects, as predicted by the Dendrophilia hypothesis. Section 4 relates Kanzi's performance to research on the acquisition of coordinate structures by human infants. Children only slowly acquire the tree structures necessary for correct comprehension of coordinate NPs. This suggests that the inter-species difference is more accurately characterized as dendrophobia on the part of Kanzi, rather than human dendrophilia. Section 5 concludes.

2. Hierarchical Structure and Artificial Grammar Learning

Fitch & Hauser (2004) tested humans and cotton-top tamarins on their ability to recognize strings conforming to one of two patterns. Each string is composed from two different types of consonant–vowel speech syllable, a and b, clearly distinguishable along several perceptual dimensions. One pattern, $(ab)^n$, consists of repeated groups of a single a syllable followed by a single b syllable. The other pattern, a^nb^n, consists of a certain number of a

syllables followed by the same number of b syllables.

Fitch & Hauser found that after familiarization to this pattern, tamarins recognized the $(ab)^n$ pattern, as indicated by an increased number of looks to stimuli inconsistent with the pattern, but did not recognize the $a^n b^n$ pattern. Adult humans, in contrast, easily recognized both patterns.[1]

This result is taken to be revealing because of differences in the expressive power of grammars required to represent these two patterns in the limit. $(ab)^n$ can be generated by a **regular grammar**, or a grammar where each rewrite rule is of the form $A \rightarrow xB$, or $A \rightarrow y$, where A and B are nonterminal nodes and x and y are terminal nodes. A regular grammar for generating $(ab)^n$ is in Figure 1.

Figure 1 also contains a tree diagram representing the structure assigned by that grammar to the string *ababab*. Although this represents a hierarchical structure of sorts, the tree diagram does not add any further information beyond that encoded in the string, in the sense that each constituent represented in the tree diagram corresponds to a final substring of *ababab*. This general property of regular grammars is related to Chomsky's (1956, 1959) claim that regular grammars cannot represent phrase structure in the general case.

No regular grammar can generate $a^n b^n$ for arbitrary n (Chomsky 1956).[2] In brief, the

[1] Several papers (Perruchet & Rey 2005, Hochmann et al. 2008, de Vries et al. 2008) have disputed this claim, on various empirical grounds. Likewise, similar experiments have reported similar results, sometimes with significant qualifications, across a range of species and stimulus types (Gentner et al. 2005, van Heijningen et al. 2009, Abe & Watanabe 2011, ten Cate & Okanoya 2012, Stobbe et al. 2012). Here, we are not concerned with the robustness or generality of the original finding, and focus instead on its relevance to the Dendrophilia hypothesis.

[2] If n is restricted to some finite set of values, $a^n b^n$ is a finite set of strings. Any finite set of strings can be generated by a regular grammar, structured as a disjunction over those

1. $S \rightarrow aT$
2. $T \rightarrow bS$
3. $T \rightarrow b$

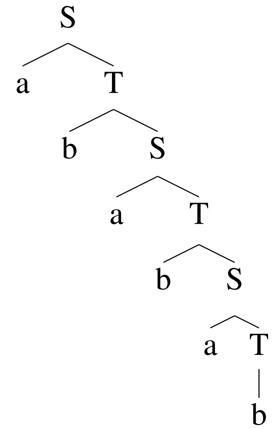

Figure 1: A regular grammar for generating $(ab)^n$, along with a tree diagram for the string *ababab* generated by the grammar.

reason is that generating $a^n b^n$ in the limit requires an arbitrary number of nonlocal dependencies pairing each a with a corresponding b. Regular grammars can handle only a finite number of nonlocal dependencies.

However, a simple **context-free grammar** can generate $a^n b^n$. In a context-free grammar, every rewrite rule in R is of the form $A \rightarrow X$, where A is a single nonterminal node and X is some string formed by concatenation of nonterminal and terminal nodes. Because the characterization of rewrite rules in context-free grammars is strictly more inclusive than that for regular grammars, regular grammars form a proper subset of context-free grammars. A context-free grammar for generating $a^n b^n$ is given in Figure 2, along with a tree diagram for *aaabbb*.

Context-free grammars are of particular interest to linguists and cognitive scientists because they can be used to represent most observed phenomena in natural language syntax (though not all — see Shieber 1985, and work such as Weir 1988 and Stabler 1997 on the more expressive **mildly context-sensitive** family of grammar formalisms). However, the

strings. Accordingly, we are only interested in formal properties of these stringsets in the limit.

1. $S \to aSb$
2. $S \to ab$

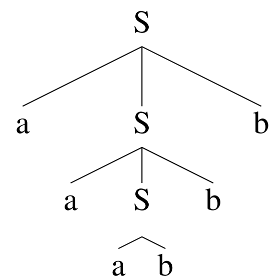

Figure 2: A context-free grammar for generating $a^n b^n$, and a tree diagram for *aaabbb*.

$a^n b^n$ grammar is far from exploiting the full expressive power of the context-free formalism. In particular, the $a^n b^n$ grammar makes use of only a single nonterminal symbol, S, while context-free models of natural language grammar use multiple nonterminal symbols to distinguish constituents of different categories. Context-free grammars with a single nonterminal symbol are equivalent to **counter grammars** (Chomsky 1959), a formalism that is strictly intermediate in expressive power between regular grammars and full context-free grammars. Because $a^n b^n$ can be generated by formalisms with sub-context-free expressive power, it is a poor stringset for diagnosing sensitivity to hierarchical structure of the type associated with context-free models of human language processing.

Similarly, $(ab)^n$ is far from exploiting the full potential of the regular grammar formalism. $(ab)^n$ is a **strictly 2-local** stringset in the terms of Rogers & Pullum (2011). A strictly 2-local stringset is one which can be recognized using a moving window two symbols long. $(ab)^n$ is strictly 2-local because every bigram in the stringset must be a member of the following set: $\{\langle START, a\rangle, \langle a, b\rangle, \langle b, a\rangle, \langle b, END\rangle\}$. Regular grammars generate a proper superset of the strictly local stringsets. For example, Figure 3 describes a regular grammar which generates strings containing a followed by b, or c followed by d. a cannot be followed by d and c cannot be followed by b. However, arbitrarily many xs can be inserted between these two paired symbols. This guarantees that the pairing of the symbols cannot be verified by any strictly local system.

1. $S \to aT$
2. $T \to xT$
3. $T \to b$
4. $S \to cU$
5. $U \to xU$
6. $U \to d$

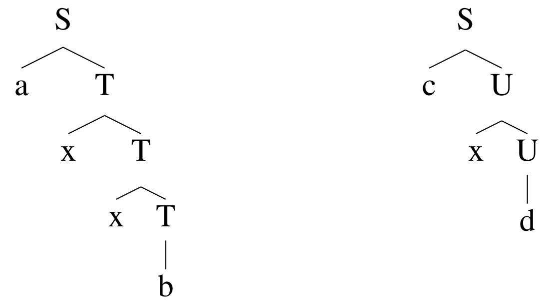

Figure 3: A regular grammar which generates a stringset which is not strictly local, and tree diagrams for two strings that it generates.

These four classes can be arranged in a strict hierarchy of expressive power, as follows: Strictly 2-local $\subset$ Regular $\subset$ Counter $\subset$ Context-free, thereby expanding the bottom half of Chomsky's (1959) hierarchy and allowing a more precise statement of the expressive power required to generate different stringsets. Many researchers (Weir 1988, Rogers & Pullum 2011, Jäger & Rogers 2012, among others) have produced still more precise characterizations of different levels of expressive power.

This is important because as our understanding of these subclasses grows, the distinction between $(ab)^n$ and a^nb^n becomes less directly relevant to the Dendrophilia hypothesis. When the two stringsets were first discussed in Chomsky (1956), only four classes of grammar (regular grammars, context-free grammars, transformational grammars, and unrestricted rewriting systems) were considered. Within that hypothesis space, a demonstration that a^nb^n is supra-regular but context-free would mean that results about learning of a^nb^n directly inform hypotheses about manipulation of context-free hierarchical phrase structure. Within a more expansive set of candidate grammar classes, such direct links cannot be drawn. As Jäger & Rogers write:

‘So what can one say about a mechanism that can learn a properly context-free pattern [like $a^n b^n$]? For one thing, it is not finite-state… Beyond that, there is very little if anything that we can determine about the nature of that information and how it is used simply from the evidence that an organism can learn the pattern.' (Jäger & Rogers 2012: 1962)

In particular, we cannot draw robust inferences about the Dendrophilia hypothesis from comparisons of learning of $a^n b^n$. To reinforce this point, we will take a brief detour into automata theory.

For each class of grammar defined above, there is a corresponding class of automata. These are useful from a cognitive perspective as automata capable of recognizing certain patterns can be thought of as minimal models of minds with the same capabilities.

Strictly local stringsets are recognized by 'moving window' automata informally characterized above — we will not be more precise about them here, but see the discussion of 'k-limited automata' in Chomsky (1963). Regular stringsets in the general case are recognized by **finite-state automata**. Figure 4 shows a diagram representing a finite-state automaton for recognizing the language $ax^*b \mid cx^*d$ from Figure 3 This diagram can be interpreted as follows: the circles represent states, and the arrows represent transitions between states. We begin in the **initial state** S and follow the arrows until we reach the **accepting state** E. Along the way, we add any symbols associated with the transition arcs to the end of the output string. The nonlocal dependencies between a and b, or c and d, are captured by the fact that the automaton can follow the loops in the intermediate states arbitrarily often.

For every context-free grammar, there is an equivalent **pushdown automaton**, consisting of a finite-state automaton augmented with memory in the form of a pushdown stack. The transitions can manipulate the top element on that stack, as well as write output symbols. A pushdown automaton for recognizing natural language data typically uses the stack

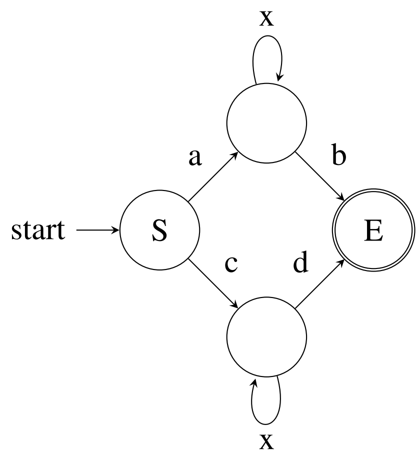

Figure 4: Finite-state automaton for recognizing the stringset $ax^*b \mid cx^*d$.

to record information about categories needed to complete a sentence. For example, an automaton processing the English sentence *The dog barks* may encounter the determiner *the*, and use the stack to record that the sentence can be completed by a noun followed by a verb phrase, information implicit in context-free rewrite rules such as $S \to NP\ VP$ and $NP \to D\ N$.

The relationship between finite-state automata and pushdown automata makes it clear that the difference between regular languages and context-free languages comes down to memory: recognizing a context-free language requires an ability to remember a 'to-do list', or a stack of nonterminal nodes which have not yet been expanded. Recognizing a regular language requires only memory of a location in a transition network. Recognizing a strictly 2-local language needs only a memorized list of accepted bigrams.

Recognizing a counter grammar requires a type of memory intermediate between finite-state automata and pushdown automata. A counter automaton for recognizing a^nb^n is in Figure 5, where $+1$ and -1 can be thought of as an abbreviation for 'add an arbitrary symbol to the stack' and 'remove a symbol from the stack', respectively. As with a pushdown automaton, the derivation terminates in an accepting state with an empty stack.

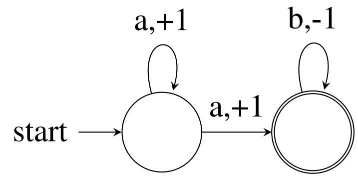

Figure 5: A counter automaton for recognizing $a^n b^n$

In other words, counter grammars are unable to differentiate among symbols written to the stack but can only manipulate the number of symbols on the stack. In contrast, a pushdown automaton's behaviour can be conditioned in different ways by different symbols on the stack. A counter automaton therefore offers a formalization of the intuition that we recognize $a^n b^n$ by counting, relating the type of memory required for counting to classes of automata both more and less expressive. To the extent that the notion of phrase structure is dependent on differentiation of phrasal categories (represented by symbols on the stack), counter automata also demonstrate quite clearly that inferences from $a^n b^n$ to hierarchical structure as found in natural language are invalid.

We should not be surprised that the full expressive power of a context-free grammar or pushdown automaton is not used in recognizing $a^n b^n$, because the hierarchical structure generated by the context-free grammar in Figure 2 is an example of **centre-embedding**, shown by Miller & Chomsky (1963) to seriously hinder sentence processing.[3] As discussed

[3] Fitch has commented in several places (e.g. Fitch & Friederici 2012, Fitch 2014) that it is important to distinguish notions such as supra-regularity, hierarchy, centre-embedding, and recursion when discussing $a^n b^n$. This advice is well taken. However, the notions are not completely unrelated. If we are to use $a^n b^n$ to inform discussions of Dendrophilia, and if a core domain in which Dendrophilia is in evidence is natural language, then we can only evaluate the Dendrophilia hypothesis by considering $a^n b^n$ *as generated by grammars like those used in natural language*. This appears to be the position of Fitch & Hauser

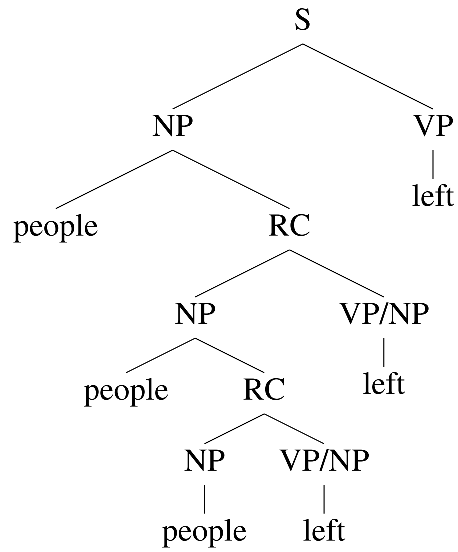

Figure 6: A tree diagram for the well-formed sentence of English *People people people left left left*.

by Rogers & Pullum (2011), the strings *people left*, *people people left left*, *people people people left left left*, and so on are all grammatical sentences of English, because *people left* can function as a reduced relative clause modifying a preceding bare plural nominal, as in *Gifts people left were donated to charity*. If *people left* can modify *gifts*, we assume it can modify *people*. Accordingly, $people_1$ $people_2$ $left_2$ $left_1$ is derived from $people_1$ $left_1$ by attaching the modifier $people_2$ $left_2$ to $people_1$.[4] Accordingly, by induction, the string $people^n$ $left^n$ is a grammatical sentence of English for $n \geq 1$, with structures as in Figure 6.

However, the grammaticality of $people^n$ $left^n$ is masked by the extreme processing difficulty caused by recursive embedding of noun phrases in the middle of larger noun phrases,

(2004), who include a tree diagram for a^nb^n showing recursive centre-embedding, and Fitch (2014), who links a^nb^n to his Dendrophilia hypothesis. Alternatively, if we see a^nb^n as merely supra-regular, then it tells us nothing about recursion or centre-embedding, but also tells us nothing obvious about the Dendrophilia hypothesis. To relate a^nb^n to phrase structure without invoking recursive centre-embedding would be incoherent.

[4]The subscripts are added purely for clarity and have no theoretical status.

and relative clauses in the middle of larger relative clauses. Although the impact of centre-embedding on processability can be mitigated by certain semantic manipulations (Crain & Steedman 1985), those manipulations involve greater differentiation among constituents at different levels of embedding. Clearly, this does not apply in the case of $\textit{people}^n\ \textit{left}^n$, where there is no such differentiation.

Our inability to recognize the well-formedness of $\textit{people}^n\ \textit{left}^n$ should raise serious doubts as to whether we use a centre-embedding context-free grammar to recognize $a^n b^n$. The existence of an analysis based on a counter grammar offers a plausible alternative, but seriously problematizes any inferences about hierarchical structures in cognition on the basis of comparisons between $(ab)^n$ and $a^n b^n$.

Related conclusions are reached by Perruchet & Rey (2005) and Lai & Poletiek (2011), among others. For example, Lai & Poletiek trained participants on stimuli constructed in such a way that the choice of *b* element was contingent on the choice of *a* element: *a* elements *be* and *bi* were paired with *b* elements *po* and *pu*, while *a* elements *de* and *di* were paired with *to* and *tu*, and so on. In other words, *bepo* and *dibepotu* were well-formed elements of this language, but *beto* and *dibepopu* were not, because an *a* element is paired with an incorrect choice of *b* element. This contrasts with the grammar in Fitch & Hauser (2004), which did not differentiate among subclasses of *a*-element and *b*-element. This manipulation yields a grammar which requires a learner to track the dependencies between *a* elements and *b* elements as represented in Figure 2, rather than simply count them. Lai & Poletiek show that humans fail to learn this strictly context-free grammar except under specific 'starting small' training regimes which include initial exposure to well-formed *ab* pairs without centre-embedding. However, Rey, Perruchet & Fagot (2012) show that under appropriate intensive training regimes (an average of 53,349 training trials per individual), baboons can succeed at formally similar tasks, and prefer to use $a^n b^n$ orders with strictly nested dependencies over other $a^n b^n$ orders. Again, then, evidence from success at learning $a^n b^n$ languages does not appear to support the Dendrophilia hypothesis directly.

In sum, comparisons between learning of $(ab)^n$ and $a^n b^n$ are unlikely to be particularly informative about the Dendrophilia hypothesis. This is perhaps not an intrinsic problem with artificial grammar learning experiments: as discussed in Rogers & Pullum (2011), Fitch & Friederici (2012), and Jäger & Rogers (2012), other stringsets may offer a more accurate picture of the cognitive resources underpinning a given species' pattern-recognition abilities. However, such work is as yet inconclusive. In this paper, I turn to a complementary source of evidence.

3. The Kanzi Corpus

This section discusses a corpus of the behavioural responses of a bonobo, Kanzi, to 660 spoken English instructions, published in Savage-Rumbaugh et al. (1993). Although the syntactic structure of these instructions is not as tightly controlled as in Fitch & Hauser's experiment, the data have four clear advantages. First, the test sentences directed at Kanzi are all well-formed sentences of English, a language with hierarchical sentence structure. It is therefore easier to relate the results to the Dendrophilia hypothesis. Second, the range of structures used in the sentences directed at Kanzi is greater than in the Fitch & Hauser experiment, providing indirect information about a wider range of representational capabilities. Third, the structural information we will probe is independent of concerns relating to centre-embedding, which removes one potential confound in interpreting the results. Finally, the sentences are interpreted by Kanzi, which provides a further source of indirect information about the structures he assigns to the sentences he hears.

3.1 Background

Kanzi acquired a high level of competence in an English-like communication system without explicit instruction (Savage-Rumbaugh et al. 1993, Savage-Rumbaugh & Lewin 1994, Savage-Rumbaugh et al. 1998). He quite quickly acquired a lexicon of several hundred items, and showed a high degree of comprehension of English utterances, either spoken or encoded on a special keyboard. The tests reported in Savage-Rumbaugh et al. (1993) are intended as a measure of Kanzi's comprehension. The instructions used are novel and unpredictable, and presented in such a way to avoid unconscious cues from the experimenter. Each item in the corpus is in the format shown in (1). First is an item number (I reproduce this in lieu of a full citation — all items can be found in Savage-Rumbaugh et al. 1993: 111–210). Next is a code for the correctness of Kanzi's response: C and C1–5 are correct, but with increasing amounts of hesitation or human intervention; PC (Partially Correct), OE (Object Error), and other codes describe a range of incorrect responses. Next, in italics, is the sentence uttered. Finally, in parentheses, is a description of Kanzi's actions and a justification of the code assigned, if necessary.

(1) 287. (C) *Kanzi, take the tomato to the colony room.* (Kanzi makes a sound like "orange"; he then takes both the tomato and the orange to the colony room.) [C is scored because it is assumed that Kanzi is announcing that he wants to take an orange and have it to eat.]

Our interest is in the distribution of 'correct' responses (coded C or C1–C5) versus incorrect responses (including PC and OE) across different syntactic structures. Savage-Rumbaugh et al. (1993: 77) give Kanzi's overall accuracy across the corpus as 71.5%, slightly higher than the 66.6% accuracy of Alia, a human infant tested on a similar set of utterances over a 6-month period starting when she was 18 months old. In Sections 3.2–3.3, we use Kanzi's general high performance to investigate comprehension strategies which he is able to use, before turning in Section 3.4 to a significant dip in performance which reveals an inability to manipulate hierarchical syntactic representations, and therefore constitutes evidence for dendrophobia in Kanzi's grammar learning.

3.2 Comprehension without Structure-sensitivity

Although the instructions directed at Kanzi are novel, they are mainly built from predictable parts: simple action descriptions such as *give*, *put*, *show*, *take*, and *get*, and a large class of nominal expressions. The nature of the objects described by those nominal expressions is a fairly reliable predictor of the way in which Kanzi is asked to interact with those objects. In the test sentences, objects are generally given or shown to animate beings, placed in receptacles, bodies of liquid, etc., and on objects that they cannot be placed in. If an object is present in Kanzi's immediate vicinity, it can be taken to a specified location. If the object is not present, it can be retrieved from a specified location. Knowing what a noun refers to therefore implies a default way of interacting with that object, and around 420/660, or 64%, of the instructions in Savage-Rumbaugh et al. (1993) ask Kanzi to act in this default way.

Success on these trials alone would be consistent with the hypothesis that Kanzi has acquired an impressive lexicon, can construe instructions as instructions, and can figure out plausible responses based on an array of default actions and the properties of the named objects and events, but that he is lacking any syntactic and/or combinatorial semantic ability. This would constitute what Anderson (2004) calls a 'semantic soup' strategy: the meanings of the individual content words are formed into a coherent action description in whatever way they fit best, without attention to any syntactic information in the signal.

However, there is evidence that Kanzi moves beyond semantic soup to a limited extent.

In certain ‘reversible’ ditransitive sentence frames, semantic soup is insufficient to determine a correct mapping from words to argument roles. For example, in (2), either action is equally plausible on the basis of the semantics of the individual lexical items alone. Sensitivity to linear order is required to associate the right object with the right argument role: the first noun encountered denotes a THEME and the second denotes a GOAL.

(2) a. 525. (C) *Put the tomato in the oil.* (Kanzi does so.)

b. 528. (C) *Put some oil in the tomato.* (Kanzi picks up the liquid Baby Magic oil and pours it in a bowl with the tomato.)

There are 43 sentences presented in such alternations in the corpus (21 pairs, with one sentence repeated, Savage-Rumbaugh et al. 1993: 95–6). Kanzi responds accurately to 33 of them (76.7%), in line with his 71.5% overall accuracy across the corpus.

This sensitivity to linear order represents a step beyond a ‘semantic soup’ interpretive strategy. However, the Dendrophilia hypothesis is concerned with a further step: the ability to make inferences about hierarchical structure from sequential information. Section 3.3 examines, and ultimately dismisses, some cases which initially appear to show sensitivity to hierarchical structure. Next, Section 3.4 shows that Kanzi fails to infer hierarchical structure when correct interpretation requires it.

3.3 Nonevidence for Structure-sensitivity

Many instructions directed at Kanzi are comprehensible without reference to phrase structure, typically by relying on a semantic soup strategy, even when they involve complex constituents from the perspective of standard English phrase structure. I illustrate this with reference to noun phrases, because (from the perspective of standard English grammar) noun phrases are the locus of most of the complexity of the utterances directed at Kanzi. Similar points hold, more trivially, of other would-be constituents.

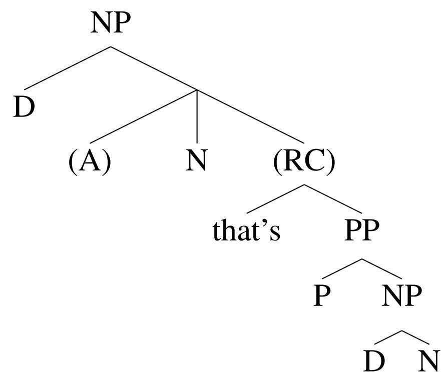

Figure 7: The maximal noun phrase structure used in instructions directed at Kanzi, from the perspective of a standard English grammar.

The maximal noun phrase structure used in the instructions directed at Kanzi is given in Figure 7. It consists of a determiner and a noun, possibly with an intervening adjective, and possibly with a following relative clause of the form *that's PP*, where *PP* describes a location. Examples with adjectives are given in (3), and with relative clauses in (4).

(3) a. 563. (C) *Show me the hot water.* (Kanzi vocalizes.) *Can you show me the hot water?* (Kanzi makes a sound like "hot," then picks up a paint brush, [and] points to the hot water with it, [. . .])

b. 564. (C) *Can you pour the ice water in the potty? Pour the ice water in the potty.* (Kanzi picks up the bowl of ice water and heads toward the potty.) E [the experimenter] says, "That's right." (Kanzi pours the ice water carefully into the potty.)

(4) a. 500. (C) *Get the lighter that's in the bedroom*. (Kanzi does so.)

b. 508. (C) *Go get the lighter that's outdoors*. (Kanzi goes to the play yard, picks up the lighter, and [. . . eventually] brings the lighter back.)

In most cases, there is only one noun in the sentence that a modifier may plausibly relate to, so a semantic soup strategy is sufficient to derive the correct association between modifier and modifiee. For instance, (3a) and (4a) could produce an accurate response even if interpreted as ‘show me water (“hot” is somehow relevant)’ and ‘get lighter (“outdoors” is somehow relevant)’.

It is possible in principle to distinguish structure-sensitive interpretation of modifiers from semantic soup, using even quite simple sentences. One way to do this would be using multiple modifiers, each modifying a different noun, as in (5). Hierarchical structure would facilitate the association of the right modifier with the right noun.

(5) a. Pour the hot water in the ice water.

b. Put the rock that’s outdoors on the TV that’s in the bedroom.

Unfortunately, such sentences are not tested in Savage-Rumbaugh et al. (1993). At present, then, there is no evidence from modification patterns to warrant the inference that Kanzi’s strategies for understanding English make reference to noun phrases as constituents.

3.4 When Structure-sensitivity is Necessary

Part of understanding the lexical semantics of a verb involves knowing how many **arguments** it takes, or how many participants are involved in the event it describes. We know that giving requires a giver, a receiver, and a thing given. Likewise, scaring requires a scarer and a scaree. There is typically a 1–1 mapping between semantic participants and noun phrases used to describe those participants.

In most natural languages, a noun phrase can be arbitrarily complex. One device for creating complex noun phrases is **NP-coordination**, which creates noun phrases of the form NP_1 *and* NP_2. These coordinated noun phrases occupy a single argumental position. Correctly interpreting a coordinate noun phrase therefore requires an ability to represent NP_1 *and* NP_2 as a constituent, whose reference is determined jointly by the component noun phrases. In examples like (6), the interpretation is that the action of fetching is applied to the object denoted by each coordinated noun phrase.

(6) Fetch [[the ball] and [the rock]]

A dendrophobic grammar without hierarchical constituent structure might include a principle mapping nouns (or object-denoting words) to verbal arguments. Sentences containing a coordinate noun phrase systematically violate such a principle, in that the number of nouns is greater than the number of argumental noun phrases. This poses an interpretive challenge: if *Fetch ball* is a well-formed instruction to Kanzi,[5] what is he to make of the extra noun in *Fetch ball rock*?

Without a representation of constituency, three interpretations of NP-coordination are likely. One might ignore the first noun, ignore the second noun, or interpret both as a coordinated object (the fourth logical possibility, ignoring both nouns, would only multiply the interpretive problems, as *fetch* would then be missing an argument). As a null hypothesis, we might expect all three actions to be equally frequent. In that case, chance behaviour would be to respond correctly $\frac{1}{3}$ of the time, and to ignore each conjunct $\frac{1}{3}$ of the time.

The number of relevant sentences in the Savage-Rumbaugh et al. corpus is small, but the patterns conform to these predictions.[6] The corpus contains 26 instructions given to Kanzi involving a coordinated NP object. Eight of these are excluded under three different criteria. First, in four cases (excluded under criterion *a* in the Appendix), Kanzi was asked

[5]There is no evidence that Kanzi interprets determiners or other functional material.

[6]My conclusions about Kanzi's accuracy, based on the descriptions given, differ substantially from those reported in Savage-Rumbaugh et al. (1993: 78). That work reports the following results: Kanzi: $\frac{7}{21}$ correct trials, Alia: $\frac{12}{21}$ correct trials. The difference between these two proportions is not significant ($p = 0.21$, 2-tailed Fisher exact test). However, there are reasons to suspect that these proportions do not fully reflect the subjects' performances. For fuller discussion of this discrepancy, see the Appendix.

to interact with two other apes (Panbanisha and Panzee, or Sherman and Austin), in such a way that it is impossible to tell whether he interpreted the instruction as referring to both apes or to just one. For example, in (7a), Panbanisha and Panzee occupy the same area, so it is unclear whether Kanzi intended to scare just Panbanisha, just Panzee, or the two of them. Similarly, in (7b), giving a single orange to both Panbanisha and Panzee is highly unlikely: he has to give it to one or the other.

(7) a. 293. (C) *Get the monster mask and go scare Panbanisha and Panzee*. [. . . Kanzi spends a while looking for the mask] (Kanzi finds the mask and goes back to the tool room with it to scare Panbanisha and Panzee [. . .])

b. 311. (C4) *Take the orange to Panban and Panzee*. [. . . Kanzi initially tries to eat the orange] Rose [. . .] says, "Kanzi, stop." (Kanzi puts the orange back down on floor.) E says, "Pick it up and take it to Panban and Panzee. Carry it to Panban and Panzee. You take it to them." [. . . Kanzi gets E to open a door for him] (Kanzi goes to the middle room door and tries to shove the orange under; Rose comes and opens the door for him, and Kanzi takes the orange in to Panban and Panzee. [. . .])

Second, in two trials (excluded under criterion *b* in the Appendix), Kanzi was already in contact with one of the two objects in question at the time of the instruction, and didn't perform any further action on that object as a result of the instruction, as in the following cases.

(8) a. 271. (C) *Show Sue the ball and the cereal*. (Kanzi picks the cereal up as E starts to talk. When he hears the sentence, he grabs the ball and tries to open the door to give E the ball. He keeps the cereal in his lap but appears to be attempting to show E the ball.) [C is scored because Kanzi picks up both

requested objects. Whenever either subject is asked to ‘show’ something, they are given the benefit of the doubt if they act on it in some way that could serve to draw E’s attention and if E can see this.]

b. 297. (C) *Show me the shot and the ball*. (Kanzi shows E the shot and looks at the ball that he is leaning on.) [C is scored because Kanzi has both items in front of E and can reasonably presume that E will see them [...]]

Finally (criterion *c* in the Appendix), in two cases, Kanzi interacted with both objects named by the coordinate noun phrase, but not in the way requested. For instance, in response to item 280, *Give me the lighter and the water*, Kanzi put the lighter in the water. This trial will be ignored below, although Kanzi’s response is incorrect in an interesting way, suggesting a default response involving a basic triadic action like putting.

Discarding these eight trials leaves 18 NP-coordination trials.[7] Of these, Kanzi ignored the first NP on 9 trials, as in (9a), ignored the second NP on 5 trials (9b), and responded correctly to 4 trials (9c), an overall accuracy of 22.2%. Compared to his overall accuracy across the corpus, this is a highly significant drop in performance (binomial test, $p = 1.1 \times 10^{-5}$).

(9) a. 428. (PC) *Give the water and the doggie to Rose*. (Kanzi picks up the dog and hands it to Rose.)

b. 526. (PC) *Give the lighter and the shoe to Rose*. (Kanzi hands Rose the lighter, then points to some food in a bowl in the array that he would like to have to eat.)

c. 281. (C) *Give me the milk and the lighter*. (Kanzi does so.)

[7] Most of these discarded trials were scored as correct in Savage-Rumbaugh et al. (1993). As documented in the Appendix, this is the basis of the discrepancy between results reported here and by Savage-Rumbaugh et al..

The same trials were presented to a human infant, Alia. Alia's accuracy across the whole corpus was slightly lower, at 66%, but her accuracy on the NP-coordination trials is indistinguishable from this baseline, at $\frac{13}{19}$, or 68.4%.[8] This suggests a species-specific, construction-specific deficit. Kanzi marginally outperforms Alia across the whole corpus, but he performs much worse than both his usual standard and the human control (Fisher exact test, $p = 0.008$), on this one construction.

Although the deficit is treated as nonsignificant in Savage-Rumbaugh et al. (1993), it is discussed as follows:

> 'The simplicity of both the semantic and the syntactic components of type 2B sentences [those involving NP-coordination] suggests that Kanzi's difficulty was perhaps due more to short-term memory limitations on the overall amount of information than to processing limitations on the information that was available to him. Indeed, it seems possible that the semantic and syntactic structure in sentences such as "Feed the doggie some milk" permitted Kanzi to go beyond the typical constraints of his short-term memory system by enabling him to process or chunk the information in a meaningful manner. By contrast, sentences such as "Give the doggie and the milk" do not engage semantic chunking strategies but rather force reliance on short-term memory alone.' (Savage-Rumbaugh et al. 1993, p. 85)

The considerations in Section 2 can help make this appeal to memory limitations precise: Kanzi's English comprehension dips precisely when successful comprehension requires reference to the kind of memory required to process hierarchical phrase structure. In other words, Kanzi suffers from dendrophobia.

[8] Again, see the Appendix for an explanation of discrepancies between the figures reported here and those in Savage-Rumbaugh et al. (1993).

4. The Acquisition of Coordinate Structures

Kanzi may exhibit dendrophobia when faced with NP-coordination, but human infants are hardly dendrophiles in this respect. Gertner & Fisher (2012) use a preferential looking paradigm to examine 21-month-olds' interpretations of sentences like (10).

(10) a. The boy is gorping the girl!

b. The boy and the girl are gorping!

c. The girl and the boy are gorping!

(10a) is a 2-noun sentence describing a transitive action. (10b–c) are 2-noun sentences which, from the perspective of an adult grammar of English, describe an intransitive action, with a single coordinated noun phrase describing the agents of that action. As described in Section 3, this mismatch between the number of nouns and the number of noun phrases severely affected Kanzi's English comprehension.

Subjects were presented with two videos in parallel, one showing a transitive action with a boy as agent and a girl as patient, and one showing coordinated intransitive actions with the boy and girl both independently acting as agents. From an adult perspective, the former matches the syntactic structure of (10a), while the latter matches the syntactic structure of (10b–c). Indeed, the infants looked more to the transitive action when presented with a transitive sentence. However, they also looked more to the transitive action when presented with (10b), where the two nouns occur in the same order as (10a). In (10c), with the order of the two nouns reversed, the subjects looked significantly less at the video of the transitive action.

Gertner & Fisher suggest that this reflects a comprehension strategy in which infants at this developmental stage rely on the number and relative order of nouns in a sentence, rather than information about noun phrases as hierarchically structured constituents. This is similar to the description of Kanzi's responses in Section 3. However, children eventually converge on a hierarchically structured representation of coordination, while Kanzi apparently never has.

The emergence of hierarchical structure is gradual: Gertner & Fisher cite evidence from Arunachalam et al. (2011) that 21-month-olds can already be biased towards a coordinate interpretation through pronominalization of the coordinated noun phrases, as in (11).

(11) A: The man and the lady are going to moop.
B: Oh yes, they are going to moop. (Arunachalam et al. 2011: 22)

By 25 months, infants tend to associate transitive and coordinated intransitive actions with transitive and coordinated sentences like those in (10) (Naigles 1990), just as 21-months had failed to. However, 28-month old infants often still associate examples like (12) with transitive actions (Hirsh-Pasek & Golinkoff 1996, Fisher 2002).

(12) Find Big Bird and Cookie Monster gorping! (Fisher 2002: 273)

This suggests that children only slowly learn to generalize a hierarchically structured representation of NP-coordination across syntactic environments: they do not automatically make maximal use of the resources at their disposal for phrase-structural representation. This is not maximally dendrophilic behaviour, in that children continue to underuse hierarchical syntactic structure, relative to adult grammars, for several months. In other words, humans are not natural dendrophiles, but Kanzi, who had not acquired the grammar of NP-coordination by age 8, is a dendrophobe.

At the same time, other species seem to be able to make some limited use of nested dependencies of the sort associated with recursively embedded constituents. For example, Rey et al. (2012) show that baboons trained on the $a^n b^n$ pattern and presented with $a_1 a_2$ stimuli, prefer to complete the pattern with the order $b_2 b_1$ than any other. As predicted by the Dendrophilia hypothesis, then, but not by the recursion-only hypothesis of Hauser et al. (2002), the inter-species differences described in this paper with respect to processing and acquisition of hierarchical structure appear to be a matter of degree: nonhuman species have some ability to pair strings with hierarchical structures, but humans do so much more readily.

5. Conclusion

The results to date from cross-species comparison of grammars such as $(ab)^n$ and $a^n b^n$ have been inconclusive. Although there is scope for further comparative investigation of pattern-recognition abilities through artificial grammar learning, inferring general cognitive capacities from limited, uninterpreted stringsets is inevitably a fraught task.

However, comparative evidence from English comprehension is broadly consistent with Fitch & Hauser's main claim of cross-species differences in recognition of hierarchical sentence structure. Kanzi interprets the string NP_1 *and* NP_2 as equivalent to NP_1 alone, NP_2 alone, or NP_1 *and* NP_2 together, seemingly at random. This is a construction in which hierarchical structure is required to arrive at a coherent interpretation of the sentence, and a major gap in Kanzi's comprehension. In contrast, Alia, a human infant, had no such problem in comprehension.

Nevertheless, children around the same age as Alia exhibit behaviour similar to Kanzi's in experiments designed to probe their comprehension of coordinate structures, and only gradually come to associate NP_1 *and* NP_2 with an adult-like phrase structure across a full range of sentence frames.

This suggests a modification to Fitch's Dendrophilia hypothesis. There may well be no special affinity between humans and trees, at least not when it comes to learning of hierarchical structures underpinning natural languages. However, humans have a capacity to learn to recognize and manipulate such hierarchical structures when sufficient evidence mandates it. Kanzi, it appears, does not, the cotton-top tamarins in Fitch & Hauser (2004) did not, and the baboons in Rey et al. (2012) only did so under intensive training. In other words, humans are not dendrophiles in this domain, but certain other well-studied species are dendrophobes.

References

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

Anderson, S. (2004). *Doctor Doolittle's Delusion*. New Haven, CT: Yale University Press.

Arunachalam, S., Escovar, E., Hansen, M., & Waxman, S. (2011). Verb learning from syntax alone at 21 months. In Danis, N., Mesh, K., & Sung, H. (Eds.), *BUCLD 35: Proceedings of the 35th Annual Boston University Conference on Language Development*, (pp. 21–24)., Somerville, MA. Cascadilla Press.

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

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

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*, 137–167.

Chomsky, N. (1963). Formal properties of grammars. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, vol. 2* (pp. 323–418). New York: Wiley.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 320–358). Cambridge: Cambridge University Press.

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

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition*, *82*, 259–278.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuoscience and comparative cognition. *Physics of Life Reviews*, *11*, 329–364.

Fitch, W. T. & Friederici, A. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B*, *367*, 1933–1955.

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

Gentner, T., Fenn, K., Margoliash, D., & Nusbaum, H. (2005). Recursive syntactic pattern learning by songbirds. *Nature*, *440*, 1204–1207.

Gertner, Y. & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*, 85–94.

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

Hirsh-Pasek, K. & Golinkoff, R. (1996). *The Origins of Grammar*. Cambridge, MA: MIT Press.

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

Jackendoff, R. & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*, 211–225.

Jäger, G. & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B*, *367*, 1956–1970.

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

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium* (pp. 112–136). New York: Wiley.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge,MA: MIT Press.

Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.

Miller, G. & Chomsky, N. (1963). Finitary models of language users. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, vol.2* (pp. 419–492). New York: Wiley.

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart and Winston.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*, 357–374.

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

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

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

Rogers, J. & Pullum, G. (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language, and Information*, *20*, 329–342.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R., Brakke, K., Williams, S., Rumbaugh, D., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, *58*, 1–252.

Savage-Rumbaugh, S. & Lewin, R. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley.

Savage-Rumbaugh, S., Shanker, S., & Taylor, T. (1998). *Apes, Language, and the Human Mind*. New York: Oxford University Press.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, *8*, 333–343.

Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical Aspects of Computational Linguistics* (pp. 68–95). Dordrecht: Springer.

Stobbe, N., Westphal-Fitch, G., Aust, U., & Fitch, W. T. (2012). Visual artificial grammar learning: Comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Philosophical Transactions of the Royal Society B*, *367*, 1995–2006.

ten Cate, C. & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B*, *367*, 1984–1994.

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

Weir, D. (1988). *Characterizing Mildly Context-Sensitive Grammar Formalisms*. PhD thesis, University of Pennsylvania.