Phrase structure grammars as indicative of uniquely human thoughts

Eran Asoulin

Sydney, NSW, Australia

A R T I C L E   I N F O

A B S T R A C T

I argue that the ability to compute phrase structure grammars is indicative of a particular kind of thought. This type of thought that is only available to cognitive systems that have access to the computations that allow the generation and interpretation of the structural descriptions of phrase structure grammars. The study of phrase structure grammars, and formal language theory in general, is thus indispensable to studies of human cognition, for it makes explicit both the unique type of human thought and the underlying mechanisms in virtue of which this thought is made possible.

1. Introduction

A revealing way in which to conceptualise the kind of thought that appears to be uniquely human is by considering the generative capacity of rule systems such as grammars and the way in which such systems can be arranged in a mathematical hierarchy of increasing generative power. One way in which to do so is the so-called Chomsky hierarchy, which was originally cast as part of the explanation of the formal properties of natural languages, but it turns out that it is also useful for understanding the formal properties of types of mental computations and hence of particular thought processes. I argue that the subset of human thought that is not shared with animals can be accounted for by the underlying computational mechanisms of language that structure this subset of human thought in a particular way. This subset of thought is unavailable to cognitive systems that do not have access to the computations underlying natural language.

Fitch and Hauser (2004) argue in regard to tamarins that they "suffer from a specific and fundamental computational limitation on their ability to spontaneously recognize or remember hierarchically organized acoustic structures" based on phrase structure grammars (Fitch and Hauser, 2004: 380). Their research suggests that nonhumans are in a sense stuck in their attempt to interpret patterns based on phrase structure grammars, for they attempt to interpret such patterns on the basis of finite state grammars, which is not possible. This makes phrase structure grammar "stimuli seem much more complex to them [nonhumans] and perhaps even unlearnable in finite time" (Fitch and Hauser, 2004: 380). Humans, on the other hand, easily learn phrase structure grammars and appear to favour input that can be interpreted as such. Phrase structure grammars form a crucial component of all human languages (allowing for recursively-embedded, hierarchically-structured strings), and the computations that allow their generation and interpretation underlie a uniquely human type of thought. What I would like to explore in this article is the inference from the ability to recognise phrase structure grammars to the

existence of particular thought processes. Fitch and Hauser only make an inference from an inability to recognise phrase structure grammars to a particular computational limitation, but I think that we can conclude more than that. If animals have a fundamental computational limitation because they cannot recognise phrase structure grammars, then it follows that cognitive systems that *can* recognise phrase structure grammars do not have this computational limitation. Now, if (as I argue below) the underlying computations necessary to generate and recognise phrase structure grammars are the same computations necessary to generate and recognise human-type thought, then it follows that any system that cannot recognise phrase structure grammars also cannot think those particular thoughts (though it can of course think other kinds of thoughts not dependent on such computations).

I should note that there is a tradition in philosophy that argues that all thought requires language (Malcolm, 1972; Davidson, 1975, 1982; Dummett, 1989; McDowell, 1994), whereas others have agreed with Fodor (1975: 56) that the "obvious (and, I should have thought, sufficient) refutation of the claim that natural languages are the medium of thought is that there are nonverbal organisms that think" (see Ryle, 1968; Slezak, 2002; de Waal, 2016). I do not want to weigh in on the debate of whether cognitive processes as understood by cognitive scientists are the same as what philosophers such as Malcolm and Davidson understand to be thought processes. Nothing thus follows from my use of *thought* or *thought process* in contrast to *cognition* or *cognitive process*. Whether we should conceive of the cognitive processes that humans share with animals as thoughts or not is orthogonal to my concerns here.

This article is structured as follows: I first discuss the details of the Chomsky hierarchy, focussing on phrase structure grammars and the structural descriptions necessary for natural language sentence generation and interpretation. I then discuss the inference from the inability to interpret certain grammars on the Chomsky hierarchy to the existence of a fundamental computational limitation. I then argue that the underlying computations necessary for phrase structure grammars are the same as those necessary for human thought (though not for the kinds of thoughts we share with animals). I argue in the final section that if this is the case, then we can make the inference from the ability to interpret phrase structure grammars to the ability to think certain types of thoughts.

The discussion of the Chomsky hierarchy in what follows is not a rehearsal or a rehashing of the debates in formal linguistics in the 1950s and 1960s. Rather, what we have here is a well understood mathematical hierarchy that has in recent research been applied in a new way to a new domain. Rogers and Hauser (2010), for example, propose a range of patterns that are based on grammars on the Chomsky hierarchy and which can be used in current and future research. They argue that experiments testing whether particular animals can interpret patterns based on particular grammars on the hierarchy offer a window into the underlying psychological mechanisms of these animals. I think that such experiments also offer a window into the underlying psychological mechanisms (hence thoughts) of humans, and help clarify the nature of human thought and the underlying mechanisms that make such thought possible.

2. The hierarchy of grammars

The generative capacity of rule systems such as grammars can be arranged in a mathematical hierarchy of increasing generative power. One way to do this was first formulated by Chomsky (1956, 1959; see also Chomsky & Schützenberger 1963).[1] Though originally cast as part of the explanation of the formal properties of natural languages, it turns out that this hierarchy is also useful for understanding the formal properties of types of mental computations and hence of particular kinds of thoughts. A grammar here is understood as a computational device that enumerates the sentences of a language: it is a set of rules that gives a recursive enumeration of the sentences belonging to the language. One requirement that a grammar must meet is that it be finite: the grammar cannot be, for example, a list of word sequences because there are potentially infinitely many of these. Hence, we investigate the structure of language by studying the finite devices (i.e., the grammars) that are capable of generating and enumerating an infinite number of sentences. It is not enough, though, to have a device that can merely generate the sentences of a language, for there is more than one way in which to generate each sentence. Knowing the way in which a sentence was generated is crucial for natural language, for each way corresponds to a different structural description and thus to a different meaning. In other words, the history of the way in which a sentence was generated is crucial for its interpretation. These two ways of understanding the generative capacity of grammars are referred to, respectively, as the grammar's weak generative capacity and strong generative capacity. A grammar is said to weakly generate a set of *sentences* but to strongly generate a set of *structural descriptions* (Chomsky, 1965). Each structural description uniquely determines a sentence, but the reverse is not the case: so one weakly generated sentence can have several strongly generated structural descriptions, each corresponding to one of the sentence's meanings.

Take the sentence *The agent saw the man with the binoculars*: there are at least two different ways in which to generate this sentence ((1) and (2) below), each of which corresponds to a different interpretation of the sentence.

(1) [s [NP The agent] [VP [VP [V saw] [NP the man] [PP with the binoculars]]].
(2) [s [NP The agent] [VP [VP [V saw] [NP the man [PP with the binoculars]]]].

---

[1] The hierarchy was a systematisation and expansion of the earlier work of, amongst others, Post (1947), Rosenbloom (1950), Bar-Hillel (1953). See Lasnik and Uriagereka (2012) for discussion.

In (1), the prepositional phrase (PP) *with the binoculars* is attached to the verb phrase (VP), thus modifying the event of seeing referred to by the verbal expression. The interpretation of (1) is thus that the agent saw the man by making use of a pair of binoculars. In (2), the same PP is now attached to the direct object noun phrase (NP), thus modifying the man referred to by that nominal expression. The interpretation of (2) is thus that the agent saw a man who was holding a pair of binoculars.

It is clear in this simple example how the way in which each sentence was generated is crucial for obtaining the correct interpretation. We need to know the derivational history of the sentence during which, say, the PP was attached to the VP thus obtaining the interpretation of (1). In other words, if we don't know the way in which the sentence was generated we cannot interpret it. Hence, a theory of language must be able to produce not only the observable corpus of sentences but "must also indicate how, in general, relevant structural information can be obtained for a particular sentence generated by a particular grammar" (Chomsky, 1959: 138). That is, the theory must specify a class of structural descriptions that provide "certain information which will facilitate and serve as the basis for an account of how *x* [a sentence] is used and understood by speakers of the language whose grammar is *f*" (Chomsky, 1959: 138). It thus follows that if a cognitive system cannot generate the required structural descriptions then it cannot access the interpretations inherent in those particular structural descriptions.

A grammar that is able to generate the recursively enumerable set of all the sentences of a language is of very limited value and interest unless the underlying principles of the construction and generation of each sentence are also provided, thus shedding light on the structural information of each sentence that is used in language production and comprehension. The different properties of different types of grammars, however, mean that not all types of grammars are able to generate the required underlying structures for natural language, and thus are unable to provide the structural information that forms the basis for the comprehension and production of sentences. The hierarchy of such grammars was originally described by Chomsky as follows:

Type 0: recursively enumerable grammars
Type 1: context-sensitive grammars
Type 2: context-free grammars
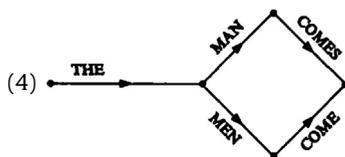Type 3: finite state grammars

Each type of grammar forms a subset of the (more complex) grammar above it, so that the set of recursively enumerable grammars includes context-sensitive grammars, and the set of context-sensitive grammars includes context-free grammars, and all of the sets in the hierarchy include finite state grammars. This relation of inclusiveness is unidirectional, so that even though the set of recursively enumerable grammars includes context-sensitive grammars, the reverse is not the case; the same is true for the other sets of grammars. It should be stressed that the Chomsky hierarchy is concerned with generalisations about strong generative capacity, with the structures underlying observable natural language sentences (see Lasnik and Uriagereka, 2012 for an excellent overview of structure in natural language).

Let us briefly see what the details of the grammars of the Chomsky hierarchy consist in. I will then argue that there are certain types of cognitive processes that are only available to cognitive systems that can compute grammars higher than type 3 (finite state grammars). That is, grammars higher up on the hierarchy, which are known collectively as phrase structure grammars, are indicative of uniquely human thoughts.

## 2.1. Finite state grammars and phrase structure grammars

A finite state grammar is the simplest possible grammar that can generate an infinite number of strings. It is usually conceived of as a machine that switches from one state to another by producing a certain symbol (say, an English word). This machine can run through a sequence of such states and produce consecutive symbols that together produce a string. A simple example of a finite state grammar can be represented by the following state diagrams reproduced from Chomsky (1957: 19).

(3a) The man comes
(3b) The men come

(4)

(5a) The old old man comes

(5b) The old old men come

(6)
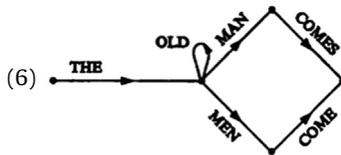
(4) is the finite state grammar that produces just (3a) and (3b), but it can be extended to allow it to produce an infinite number of strings by adding a closed loop as in (6); it is possible to add any number of closed loops. A string is produced by following a particular path of the arrows from the left to the final point on the right, and the further along the arrow path we go the more restrictions are placed that limit the possible words that can be added to the sequence.

The grammars that produce sentences and thus languages in this manner, known mathematically as finite state Markov processes, form the lowest level of the Chomsky hierarchy and they are contrasted with phrase structure grammars, which form the top three levels of the hierarchy. If the aim is the construction of a grammar of a natural language, then it can be mathematically shown that it is impossible, not just difficult, to construct a finite state grammar that will produce all and only the grammatical sentences of English (Chomsky, 1957). This applies to all natural languages, not just English. That is, due to the computational limitations inherent in finite state grammars, they cannot generate the required structures of what may at first seem like relatively simple languages. For example, a language that includes all and only sentences of n occurrences of a followed by n occurrences of b (i.e., $a^n b^n$, which generates ab, $aabb$, $aaabbb$ etc.) cannot be generated by finite state grammars because the a's and b's in question are not consecutive but rather are embedded in other strings. In other words, a finite state grammar has no way of keeping track of the structural dependency between the a's and b's, it cannot keep track of how many a's have been generated so that it can then go on and generate that exact same number of b's. These sorts of dependencies are essential for natural language sentence production and comprehension. The language described by $a^n b^n$ is a simple phrase structure grammar: specifically, it is a context-free grammar that is one up on the Chomsky hierarchy from finite state grammars. The difference in generative capacity between finite state grammars and phrase structure grammars in general is that the latter allow for self-embedding. That is, all phrase structure grammars have the ability to generate recursively embedded and hierarchically structured strings.

Sentences of natural language are objects structured in a specific way, but the strings that finite state grammars generate have the wrong sort of structure in regard to natural language. In other words, the structures generated by phrase structure grammars are essential for natural language but cannot be generated by finite state grammars. This is not a mere working assumption but rather a result of formal linguists and logicians who have shown that the mathematical properties of finite state grammars are not adequate for natural language (Chomsky and Miller, 1958). Another way to put the matter is as follows. Suppose that a finite state grammar could produce, say, the extensional set of all English sentences (suppose we could draw the state diagrams of all English sentences). That would not be sufficient because what matters for natural language interpretation are the structural descriptions in virtue of which meaning is made possible. If a cognitive system cannot generate the underlying dependencies between elements of sentences then the interpretations that such dependencies specify are unavailable to it.

Before continuing, a number of clarificatory notes are in order. There are two ways in which one can understand the term *phrase structure grammars*. On the one hand, it is used to refer to all the levels on the Chomsky hierarchy above finite state grammars. So, for example, in their supporting materials for their research, Fitch and Hauser (2004) explicitly state that they understand phrase structure grammars to include types 0, 1, and 2 on the Chomsky hierarchy. In this sense the term *phrase structure grammars* refers to a class of grammars within which there is a spectrum of generative power. And so it is usual to speak of, say, a context-free phrase structure grammar. On the other hand, the term has also been used to refer to the specific phrase structure rules that Chomsky (1957) showed are insufficient to capture the grammar of English. It is important to keep these two senses of the term separate. The particular phrase structure grammar rules that Chomsky (1957) discussed are of course not sufficient to capture the complexity of natural language syntax, but this is a different point to the claim that phrase structure grammars (as a set of grammars including types 0, 1, and 2 on the Chomsky hierarchy) are necessary for the generation of natural language sentences.

There was a decades-long debate about which phrase structure grammar (context-free, context-sensitive, or recursively enumerable) is best for capturing natural language syntax. Chomsky (1957) argued early on that English cannot be captured by finite state grammars because the latter cannot allow for structures with multiple long-distance dependencies and

recursive sentence structures. Moreover, he argued that English cannot be captured by phrase structure descriptions with context-free rules either because certain syntactic phenomena (such as sentence transformations from active to passive) are beyond their formal capabilities. He concluded that natural language grammars must be type 0 grammars, which he termed transformational grammars. However, subsequent work has shown that transformational grammars are in fact too powerful and require a set of cumbersome constraints if they are to work for natural languages (see Jäger and Rogers, 2012 for a review). This means that natural languages must be lower on the hierarchy. As Fitch and Friederici (2012: 1938) remark, "it is now clear that certain phenomena of natural language require context-sensitive grammars, and most researchers in this field now agree that human languages require 'mildly context-sensitive' grammars (MSCGs): grammars whose power is just a bit beyond those capturable by a context-free grammar". The current consensus is that natural languages are mildly context-sensitive, falling somewhere between type 2 and type 1 grammars on the hierarchy (Pullum and Gazdar, 1982; Joshi, 1985; Stabler, 2004). So natural languages are phrase structure grammars in the above first sense (namely, mildly context-sensitive grammars) but in the second sense they cannot be captured by the specific set of phrase structure grammar rules that Chomsky (1957) discussed.

Henceforth I follow Fitch and Hauser (2004) in using the term *phrase structure grammars* as denoting the class of grammars above finite state grammars on the Chomsky hierarchy because, like them, I want to avoid the technical debates about the exact level on the hierarchy that natural language should be couched. This is because the main point in my discussion here is that animals cannot compute structures belonging to phrase structure grammars as a class at all. The use of the term in this way is consistent with the way Chomsky himself used the term in his early publications. To put the matter differently, the argument in regard to where natural language falls on the Chomsky hierarchy is moot in the context of this article because animals cannot even process phrase structure grammars. Their cognitive system does not allow them to compute any structures generated by even the least computationally powerful phrase structure grammar.[2]

Finally, it should be noted that the Chomsky hierarchy is theory-neutral in the sense that it is a formal mathematical hierarchy that does not depend on generative syntax. The latter's conception of phrase structure theory has changed from its earliest incarnation, the main stages to the present include X-bar Theory (Chomsky, 1970; Jackendoff, 1977), Government and Binding Theory (Chomsky, 1981) and the current theory of Minimalist syntax (Chomsky, 1995). Other approaches include Generalized Phrase Structure Grammar (GPSG) (Gazdar et al., 1985) and Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994). Though as Fukui (2001) notes, the various approaches to phrase structure theory have begun to converge and to fruitfully influence each other. For example, "given the foremost importance of features in the theory of phrase structure (and in the Minimalist program generally), the explicit mechanisms of feature systems developed in other approaches (e.g. in the GPSG/HPSG traditions) may well have an important impact on further development of the bare theory [of Minimalist syntax]" (Fukui, 2001: 403).

2.2. Where on the hierarchy does animal cognition fall?

There are a number of recent studies looking at whether animals can grasp finite state grammars and phrase structure grammars (Yip, 2006; Hilliard and White, 2009; Bolhuis et al., 2010; Abe and Watanabe, 2011; Berwick et al., 2011a,b). These studies use the artificial grammar learning paradigm, which makes use of formal language theory to assess what types of patterns animals can detect in artificially prepared string sets (see Fitch et al., 2012 for an overview). As part of their ongoing work on nonhuman primates, Fitch and Hauser (2004) created two simple grammars, one a finite state grammar and the other a phrase structure grammar, and tested cotton-top tamarins to determine whether they could master either of them. The newly created grammars were used to generate meaningless auditory strings consisting of consonant-vowel speech syllables. Each grammar consisted of two classes of syllables (*a* and *b*), each of which contained eight different consonant-vowel combinations. The finite state grammar was of the form $(ab)^n$, in which a random *a* syllable was always followed by a random *b* syllable, and then such pairs were repeated *n* times. An example of this stimulus presented to the tamarins is *ab ab* (containing the syllables *no li | ba pa*) and *ab ab ab* (containing the syllables *la pa | wu mo | no li*). The phrase structure grammar was of the form $a^n b^n$, in which *n* sequential *a* syllables must be followed by *n* sequential *b* syllables. An example of this stimulus is *aa bb* (containing the syllables *yo la | pa do*) and *aaa bbb* (containing the syllables *ba la tu | li pa ka*) (Fitch and Hauser, 2004: 378). The two grammars were matched for length, composition, loudness, and other acoustic features and were in general designed to differ in nothing but their capacity to generate hierarchical phrase structures (Fitch and Hauser, 2004: 378). That is, the experiment was set up in such a way as to ensure that the only difference between the grammars was that one is a finite state grammar and that the other, which is a phrase structure grammar, cannot in principle be approximated by a finite state grammar.

In order to learn the $a^n b^n$ phrase structure grammar, which is easy for humans to do, one needs more computational power than a finite state grammar can provide. That is, one needs to be able to keep track of the long-distance structural dependencies between the groups formed by the *a*'s and *b*'s. Fitch and Hauser (2004) showed that tamarins easily mastered the finite state grammar, displaying the ability to discover, without explicit training, the rules governing patterns in auditory stimuli (see also Saffran et al., 2008; Endress et al., 2009; Locurto et al., 2015). In contrast to this ability, the tamarins were unable to master the phrase structure grammar, and this failure "cannot be due to extraneous factors such as stimulus length,

---

[2] I wish to thank an anonymous reviewer for alerting me to the need to stress this point.

loudness, or other acoustic factors; inability to perceive the A and B classes; or differences in exposure, testing, or evaluation procedures, all of which were consistent between the two grammars" (Fitch and Hauser, 2004: 379). Indeed, the results of this experiment (as well as the others cited in this section) are consistent with the explanation that the tamarins attempted to parse the phrase structure grammar presented to them by building a finite state grammar based on simple translational probabilities. But this is not possible because a finite state grammar does not possess the right computational machinery for the task. In other words, the cotton-top tamarins were presented with a finite state grammar and a phrase structure grammar that were equivalent in every respect except for a particular type of hierarchical structure; in contrast to their ability to spontaneously learn the finite state grammar, they were unable to spontaneously learn the phrase structure grammar that is indicative of the sort of hierarchical structure that is essential for natural language syntax.

Note that the claim here is not that the totality of an animal's cognitive processes are as representationally powerful as the most powerful grammar that the animal can learn. I don't think one can draw such general conclusions. Moreover, I am not arguing that humans are superior, nor that a purported superiority of humans is due to their ability to master grammars higher up on the Chomsky hierarchy. Nothing in my argument precludes some animals having more powerful representational resources in other domains. Rather, the claim here is that humans have a cognitive process not shared with other animals, and that what makes this possible is the ability to reflexively master phrase structure grammars. The ability of non-human animals to master finite state grammars, and their inability to master phrase structure grammars, has been investigated in birds, marine animals, and other animals (see Hilliard and White, 2009 for a review of the evidence). All such studies point to the conclusion that non-humans animals, as Fitch and Hauser put it in regard to tamarins, "suffer from a specific and fundamental computational limitation on their ability to spontaneously recognize or remember hierarchically organized acoustic structures" (Fitch and Hauser, 2004: 380). In other words, "nonhumans are 'stuck' trying to interpret PSG-generated stimuli at the FSG level [… thus making] PSG stimuli seem much more complex to them and perhaps even unlearnable in finite time" (Fitch and Hauser, 2004: 380). Humans, on the other hand, easily learn phrase structure grammars and appear to favour input that can be interpreted as such.[3] Phrase structure grammars form a crucial component of all human languages (allowing for recursively-embedded, hierarchically-structured strings), and the mental computations that allow their use and interpretation underlie a specific type of cognitive process that appears to be only available to humans and that is discussed below.

So there are two inferences here that need to be discussed and justified. One is the inference from the inability to interpret certain grammars on the Chomsky hierarchy to the existence of a fundamental computational limitation. The second inference has as its premise that the underlying computations necessary for phrase structure grammar are the same as those necessary for human thought. If this is the case, then we can infer from the ability to interpret phrase structure grammars to the ability to think certain types of thoughts. I discuss each of these in turn below.

3. The inability to interpret certain grammars as evidence of a fundamental computational limitation

Let us look at some examples of what the inference from the inability to interpret certain grammars on the Chomsky hierarchy to the existence of a fundamental computational limitation amounts to. In their discussion and review of the use of formal language theory in studies of the ability of humans and animals to recognise grammars on the Chomsky hierarchy, Rogers and Hauser (2010) remark that the aim of this line of research is to "set up a range of patterns or stringsets that enable more systematic explorations and discoveries of the underlying psychological mechanisms" (Rogers and Hauser, 2010: 215). The task of a subject in these experiments was to infer the pattern of a large, possibly infinite, stringset from a small sample that contained only positive examples (recall the $(ab)^n$ finite state grammar discussed above). Since subjects clearly cannot "extract patterns that are more complicated than they are able to distinguish", Rogers and Hauser argue that "the stringset they [the subjects] arrive at is an indicator of the capacity of the cognitive machinery they can bring to bear on the task" (Rogers and Hauser, 2010: 215). Knowing the types of grammar (or, more generally, the types of rule-based patterns) that a subject can and cannot recognise reveals a great deal about the cognitive mechanisms of that subject. In other words, if a computation is necessary for recognising a certain pattern, and if we have good evidence that a cognitive system cannot recognise this pattern, then we have good evidence that the cognitive system is not capable of the underlying computation.

Fitch and Friederici (2012) show that in order to recognise patterns based on the $a^n b^n$ phrase structure grammar a cognitive system must be able to do more than simply count, which many vertebrates are capable of. That is, in order to recognise $a^n b^n$, a cognitive system must be able to count and compare across phrases. At the very least the system must be able to tally the number of a's and store that number, then tally the number of b's and store that number, and then compare the two tallied numbers (this is the so-called count-and-compare computation). If the system can represent the exact number of a's and b's, and if these values are not bounded a priori to a fixed number, then the pattern corresponding to the grammar cannot be recognised or generated by any algorithm on the finite state grammar level of the Chomsky hierarchy (i.e., at least a context-free grammar is needed). Fitch and Friederici remark that "[t]he evidence from animals, thus far, suggests that *this* computation [of count-and-compare], unlike counting, is difficult or impossible for most tested non-human species" and that

---

[3] Indeed, even though prosodic or statistical cues can aid the human listener, hierarchical linguistic structures in speech do not have boundaries that are clearly defined, and so the structures must be constructed internally on the basis of the listener's grammatical knowledge. See Berwick et al. (2011b), Ding et al. (2016).

this "failure seems very relevant to any detailed analysis of the computational capabilities of different species' brains" (Fitch and Friederici, 2012: 1943, emphasis in original). That is, the count-and-compare computation is needed in order to recognise the $a^n b^n$ grammar, and so we can conclude that if a cognitive system cannot recognise a pattern based on the $a^n b^n$ grammar then it cannot implement the count-and-compare computation. Moreover, the structural information and interpretations inherent in the $a^n b^n$ grammar is of course unavailable to such a system. Thus, we can shed light on the nature of the cognitive processes of an animal by investigating the types of grammars that they can recognise.

To put the matter another way, the following three steps embody the first inference:

(1) A particular type of pattern can only be generated by, say, the $a^n b^n$ phrase structure grammar. This is a result of formal mathematics and formal language theory. No amount of time computing the pattern using a finite state grammar will yield the required result.
(2) A certain computational ability is necessary in order to generate or recognise a particular pattern (say the count-and-compare computation necessary for recognising the $a^n b^n$ phrase structure grammar).
(3) We can thus conclude that if a cognitive system cannot recognise a particular pattern then it lacks the underlying computational ability that is necessary to recognise this pattern.

I should note that the conclusion is an empirical question, not a matter of deductive logic, for it is conceivable that the animals tested in the experiments cited above *do* have access to the computations in question but did not display them in the experiments. It is conceivable, in other words, that future experiments might show that animals do have the required computational ability. The main point I wish to highlight here, however, is that the structure of the above inference is sound. That is, given that the premises are right, we can infer the conclusion. But what makes the premises right or wrong is an empirical question. Whether animals can recognise patterns based on phrase structure grammars is an empirical question to which the studies cited above have answered negatively. But even if it could be shown that animals *can* recognise patterns based on phrase structure grammars, the structure and validity of the inference will remain. This is an inference to the best explanation: given that a cognitive system cannot recognise patterns based on phrase structure grammars, our best explanation for this fact is that it does not have access to the underlying computations that are necessary for their recognition and generation.

Another seemingly simple grammar (from the human perspective) that is nevertheless very difficult or perhaps impossible for some animals to learn even after much training is the mirror grammar ww^R, where w represents any string and R indicates that the string is to be reversed. A mirror grammar defined over a and b will generate strings such as *abba*, *baab*, *babbab*, in which the right half mirrors the left half. The computations inherent in the ww^R grammar can only be generated and recognised by context-free grammars. Such grammars are well suited to examining the ability of cognitive systems to recognise pattern-based centre-embedded phrases and bilateral symmetry in general. Stobbe et al. (2012), for example, tested such a grammar on both pigeons and kea and showed that all the tested animal subjects failed to learn discriminations based upon symmetry. That is, the birds could not see the patterns in the world that were based upon that sort of symmetry because they do not have access to the computations in virtue of which such patterns can be generated and recognised. Stobbe et al. argue that the birds' failure to learn these discriminations can be understood in terms of what grammars on the Chomsky hierarchy the birds can recognise. When presented with the stimuli the birds acquired the wrong pattern rule, but the rule they acquired was consistent with a cognitive system attempting to use computational abilities on the finite state grammar level to recognise a pattern based on a phrase structure grammar. That is, the birds detected some local dependencies in the patterns that are capturable by finite state grammars, but that was not enough in order to solve the generalisation task, for the intended rule was on the phrase structure grammar level.

Stobbe et al.'s inference from the inability of the birds to interpret certain grammars on the Chomsky hierarchy to the existence of a fundamental computational limitation is independently verified by other studies using different methods. Swaddle and Ruff (2004) showed that starlings were not able to discriminate symmetry from asymmetry. They presented the birds with symmetric and asymmetric images that resemble the complex dot patterns commonly seen on the chest plumage of starlings, but following extended training sessions the birds were not able to discriminate between the two. Swaddle and Ruff argue that although "complex dot patterns (similar to those studies here) are common in nature, it appears unlikely that starlings (and perhaps other birds) could use number asymmetry in such traits as a direct cue to mediate behavioural interactions" (Swaddle and Ruff, 2004: 29). The experiments of Huber et al. (1999) also concluded that pigeons do not classify stimuli on the basis of an underlying symmetry rule. Since the birds in these experiments could not process the patterns based upon phrase structure grammars, there is a real sense in which they do not see the world in the same way as an animal with a cognitive system that can recognise patterns based on phrase structure grammars. Research of this kind shows that figuring out what kinds of patterns on the Chomsky hierarchy a cognitive system can recognise tells us a great deal about the psychological process (and the underlying computations) of this system.

Considering the results of the above-mentioned research, Fitch and Friederici (2012) conclude that "[i]f there is a fundamental computational restriction that prevents most species from accessing even bilateral symmetry or 'count-and-compare' strategies, this is surely relevant to these species' inability to acquire the syntax of natural language, which by all accounts require supra-regular [context-free] capabilities of at least this level of computational power" (Fitch and Friederici, 2012: 1944). That is, the inability to recognise the mirror grammar ww^R shows the inability to recognise pattern-based centre-

embedded phrases. Moreover, the structural information and interpretations inherent in the mirror grammar ww^R include bilateral symmetry, and so cognitive systems that cannot compute the mirror grammar ww^R do not have access to such interpretations: they cannot construe the world in that way. In other words, the patterns (including the dependencies and structure therein) that a cognitive system can recognise embody a particular perspective on the world; the patterns the system is able to recognise reveals the way in which the system sees the world, the way in which it understands and interacts with the world.

4. The underlying computations of human thought

The previous section detailed the inference from the inability to interpret certain grammars on the Chomsky hierarchy to the existence of a fundamental computational limitation. The research cited was mostly negative in that it inferred a computational limitation from a particular inability to recognise a pattern. The second inference I would like to discuss, however, is a positive one: what follows if a particular cognitive system *can* recognise phrase structure grammars? The premise of this inference is that the underlying computations necessary for phrase structure grammar are the same as those necessary for a particular subset of human thought. I will first discuss this premise, and then argue that if this is the case then we can infer from the ability to interpret phrase structure grammars to the ability to think certain types of thoughts.

Recall that a theory of language needs to specify a class of structural descriptions that provides the necessary information for the production and comprehension of natural language sentences. The way in which this class is generated is crucial, for a grammar that generates the observable corpus of sentences is of little value unless it provides, via structural descriptions, the required structural information in the right form for interpretation. Now, if a system cannot compute the class of grammars that correctly generates the required structural descriptions for natural language then it follows that it cannot understand the interpretations inherent in these structural descriptions. On the assumption that one of the uses of language is to externalise internally constructed thoughts, we can conclude that if a system cannot compute the structural descriptions of natural language then it cannot think the particular thoughts that can only be externalised using those structural descriptions. In other words, if a cognitive system cannot compute a particular class of grammars then anything expressible only within that class is unavailable to it. There are different types of thought that humans can construct internally that can be externalised via language, many of these we share with other animals. But the type that is unique to humans has its foundation in language itself. That is, the underlying mechanisms of language (via the construction of structural descriptions) do not merely express pre-formed thoughts but also allow humans to think particular types of thoughts that are unavailable to beings who do not have these computational mechanisms. Language is thus an instrument of thought (Burton-Roberts, 2011; Chomsky, 2013; Hinzen, 2013; Asoulin, 2016).

The particular type of structural relations that are essential for the generation and recognition of sentences, and which allow for recursively embedded and hierarchically structured strings, cannot be generated by finite state grammars. Finite state grammars can generate some types of structured and hierarchical strings, but these structures are not of the right type in regard to the structural descriptions essential for natural language interpretation. The nature of the kind of thought that requires the ability to recognise and generate phrase structure grammars can be explicated in a number of ways. One way is to consider the distinction between noun phrases (NPs) and sentences. Every human language makes a distinction between NPs and sentences, which is "a distinction between, on the one hand, what is referred to and, on the other hand, assertions made (or questions asked) about what is referred to" (Carstairs-McCarthy, 2000: 250). Compare, for example, the sentence *John arrived yesterday* with the NPs *John* or *the man who lives in Sydney.* Carstairs-McCarthy (1999, 2000) argues that the NP/ sentence distinction "permit[s] a more precise characterisation of the kinds of thing that could and could not be said with a syntax that conforms rigidly to a syllabic model [i.e., one that predates modern human natural language syntax]" (Carstairs-McCarthy, 2000: 254). Every human language makes such a distinction and every human cognitive process that is not shared with animals also embodies such a distinction. Jackendoff (2002: 327) notes that the parallelism between NPs and sentences "lies in the parallelism between conceptualized objects and conceptualized situations." The upshot is that without the ability to construct natural language sentences, access to the particular conceptualised situations that these sentences encode is blocked. In other words, a cognitive system so impeded cannot conceptualise the world in that way, it cannot think about the world in that specific way. Thus, if the generative power of finite state grammars is insufficient for the generation of the sort of structural dependencies (such as self-embedding) that are necessary for natural language sentences, and if such sentences provide access to specific types of conceptualised situations, then it follows that cognitive systems without the ability to do the former have no access to the latter.

Let us consider a specific example, taken from Hinzen (2006), of the implications of humans uniquely possessing the NP/ sentence distinction as a structural design feature of our cognitive system. Consider a glass of water on a table in front of you. You can refer to the glass by using an NP, and you can refer to the water by using an NP. But when you say *This glass contains water* what are you referring to? "Clearly you are not referring to the glass," says Hinzen, "for here you would use 'This glass'; not the water, for then you would use 'this water', or 'the water in the glass'; [and] not the relation of containing" (Hinzen, 2006: 145). There is a puzzle here, for thinking that *this glass contains water* is not the same as thinking *this glass*, and then thinking *contains*, and then thinking *water*. In other words, "a sequence of ideas does not yet constitute a complex idea. But there *are* complex ideas, over and above simple ones. Therefore, thinking a complex idea (such as a sentence) is not thinking a

sequence of ideas" (Hinzen, 2006: 145, emphasis in original).[4] Cognitive systems that have the NP/sentence distinction differ markedly from cognitive systems that do not, for a sentence is not a mere list of ideas or concepts, and to capture the specific kind of conceptual content sentences provide we need to access not only the constituents of the sentence but also the specific structural relations holding between these constituents (see also Hinzen, 2012). Phrase structure grammars of the type discussed above provide humans (and as far as we know only humans) with the ability to generate and interpret these structural relations. Note that this is not a matter of complexity, for complex NPs are possible without giving rise to sentences.

Recent empirical research also supports this conclusion. de Villiers (2014: 113), for example, lays out the "arguments and evidence that I-language makes possible some concepts that might not exist in a non-linguistic creature." Such concepts include "the concept of the class of events describable by a three term reversible sentence, namely one in which the hierarchical arrangement of arguments defines the agency, namely, the direction of action." de Villiers's lab tested whether infants could learn to form concepts of three term reversible, transitive events like *boy chasing dog* prior to language. They found that children "were much better at re-enacting the event when they heard the event described – the Language condition – than in the 'No-language' condition where they received no linguistic scaffolding" (de Villiers, 2014: 105). While more empirical work is needed, the results are "certainly compatible with the idea that language provides an assist in holding onto a reversible transitive event to reproduce it with new exemplars." In other words, "the claim is that I-language allows humans to see resemblances among events that are similar in only an abstract way mediated by their linguistic description, such as the hierarchical form <boy chase dog> that unites the class of events in which a boy chases a dog, not vice- versa" (de Villiers, 2014: 100).

The claim here is that only humans can recognise the event *boy chasing dog* "because it is a structured syntactic object whose arguments are in a certain specified relation to one another: it is not enough that the elements BOY, CHASE and DOG be present in the scene" (de Villiers, 2014: 103). Moreover, the underlying mechanisms of language allow humans to construct expressions that are constituted of variables, not just particulars. That is, "[j]ust as BOY does not refer to a specific boy but instead to the set of all boys, the composed form BOY CHASE DOG can potentially refer to an infinity of possible events in which a boy is chasing a dog. This linguistic structure thus *brings similarity* to a set of events in the world that had no unity before" (de Villiers, 2014: 103, emphasis in original). de Villiers admits that this is a strong claim, but that it is also empirically testable. Her lab has provided good preliminary evidence from children and adults that support this claim (see also Dessalegn and Landau, 2008).

Another way to understand the kind of thought that requires the ability to recognise phrase structure grammars is given in Jackendoff and Wittenberg (2014), who investigate "how much and what kind of thought can be expressed without resorting to the tools provided by fully complex syntax?" and "What can semantic structure alone accomplish before syntax is needed to regiment the message?" (Jackendoff and Wittenberg, 2014: 65). They develop an alternative to the Chomsky hierarchy and propose a model of successively more complex sound-to-meaning mappings. According to their model, there comes a point in the hierarchy below which less complex grammars are a direct mapping from sound to meaning, leaving no need for a syntactic level. Syntax, they argue, is not needed in every language and language-like system for it "is one step removed from meaning: it is a formal system that abstracts away from semantic categories, that labels constituents in terms of abstract categories such as nouns and verbs, and that imposes formal structure on them" (Jackendoff and Wittenberg, 2014: 67). The point in the hierarchy at which syntax is needed, they argue, is when the required strings necessitate structural relations that cannot be generated by finite state grammars and can only be generated by at least context-free grammars.

According to Jackendoff and Wittenberg (2014: 68ff.), the simplest conceivable grammar is a one-word grammar, in which utterances are restricted to a single word. The next simplest grammar is a two-word grammar, which leads directly to a concatenation grammar that allows utterances to consist of word strings of arbitrary length. As Jackendoff and Wittenberg remark, interpretation in these three grammars is conditioned by semantic distinctions and not by syntax. In other words, in these (finite state) grammars "there is no real syntax in the usual sense, as defined by syntactic categories and syntactic markers of these categories" (Jackendoff and Wittenberg, 2014: 69). Now, when we require the grouping of words into phrases (and then the grouping of phrases into larger phrases), the need for syntax arises. A simple phrase grammar in the Jackendoff and Wittenberg model allows words to be grouped into phrases, and such grammars are among the simplest possible context-free grammars. The difference between a concatenation grammar and a phrase grammar is their interface rules, which link utterances produced by these grammars with meanings. The simplest interface rule is that used in a one-word grammar in which the meaning of the word equals the meaning of the utterance. There is thus no difference between an utterance and a word here, and so the interface rule reduces to a one-to-one mapping between phonology and semantics (between a particular sound and a particular meaning). Jackendoff and Wittenberg speculate that such simple interface rules are necessary for most primate calls.

In order to add more expressive power, however, a syntax is needed to augment the phonology-to-semantics mapping. So, for example, scaling up to more syntactically complex grammars we have one-phrase utterances that link a particular phrase to a particular meaning, which in turn is linked to a particular sound. This allows the grammar to express meanings that move beyond the meanings of the individual words. Such more complex grammars can combine the meanings of the parts via the syntax to create new meanings that are not a direct mapping from the meanings of the parts.

---

[4] Hinzen is here paraphrasing Fodor (2003: 92).

The creation of new meanings via syntactic operations is in line with the work of Pietroski (2008, 2010, 2018) in internalist semantics (see also Hornstein, 2009), the leading idea of which is that "in the course of language acquisition, humans use available concepts to *introduce* formally new concepts that can be fetched via lexical items and combined via certain operations that are invoked by the human faculty of language" (Pietroski, 2010: 247, emphasis in original). That is, meanings are (internal, and unconscious) instructions for how to access and assemble concepts of a special sort. Meaning is here understood not in an extensional sense but rather in terms of the cognitive resources (the computational procedures) that humans deploy in *generating* the meanings. The meaning of a structured concept depends on its primitive elements and on the way in which they are combined. But not any combination is possible: there is a syntax that determines how concepts can (and cannot) be combined. As we saw above, there are certain syntactic relations (such as the dependencies in phrase structure grammars) that can only be generated above a certain level in the Chomsky hierarchy. It follows, then, that a cognitive system that lacks the ability to generate and interpret phrase structure grammars has no access to the meanings that are only generable via these grammars.

Human language syntax allows us to take already existing concepts and combine them to produce distinctively new concepts that we are then able to combine again to form linguistic expressions. This process of lexicalisation and concatenation is part of the explanation of the creative aspect of language use (Chomsky, 1966; McGilvray, 2001, 2005; Asoulin, 2013). According to Pietroski's Conjunctivist account of linguistic composition (Hornstein and Pietroski, 2009; Pietroski, 2018), lexicalisation is not a process in which a previously available concept is merely labelled using a lexical item that inherits its content from the concept itself. Rather lexicalisation is a device for accessing previously available concepts which become lexical items that are used as input to linguistic operations that combine the lexical items in specific ways to introduce new formally distinct concepts. Another way to put the matter is as follows. Humans possess a great variety of pre-lexical mental representations (many of which we share with other animals). On the Conjunctivist account, these pre-lexical mental representations are linked via human syntax to formally distinct but analytically similar concepts. The latter are sometimes referred to as I-concepts (Jackendoff, 1989, 1990). Thus, "the repertoire of I-concepts expressed by sentences cannot be mentally encoded as a list, but must be characterized in terms of a finite set of mental primitives and a finite set of principles of mental combination that collectively describe the set of possible I-concepts expressed by sentences" (Jackendoff, 1990: 9). I-concepts, then, are a uniquely human *subset of concepts* that humans can use to think about the world. If lexical meanings are understood to be instructions to fetch concepts, then phrases are understood to be instructions to combine these fetched concepts in specific ways. The internalist semantics claim, then, is that understanding a linguistic expression is a matter of (unconsciously) recognising that that expression is an instruction to construct concepts of a special kind (see also Asoulin forthcoming).

As Fodor (1975) famously argued, there are parallels between the structure of language and the structure of thought. That is, the way in which the constituents of sentences are related in a systematic way mirrors the way in which the constituents of thought are related in a systematic way. Fodor argued that the reason for this is that language and thought both employ a generative procedure that allows the creation of an unbounded set of structured expressions. This procedure is a computational mechanism that allows humans to produce recursive and hierarchically structured expressions indicative of phrase structure grammars. Though note again that such a procedure of course cannot be responsible for all of our thought processes, for much of what we share with animal cognition is clearly rich and complex but does not involve language nor its underlying mechanisms. I argued above that one way in which to determine the extent to which conceptual thought is influenced or determined by the underlying computational procedures of language is by looking at where on the Chomsky hierarchy a particular cognitive system falls.

I should note that the usage-based approach to linguistics, which includes the various strands of functional and cognitive linguistics, rejects the claim that language is an independent cognitive faculty (Langacker, 1987; Tomasello, 2003, 2008; Bybee, 2010; Christiansen and Chater, 2015). Their claim is that the underlying cognitive mechanisms of language are not language-specific and that language acquisition is due to domain-general mechanisms found elsewhere in human cognition. This is an ongoing debate, and the work of, amongst others, Berwick et al. (2011a,b) and Adger and Svenonius (2015) specifically addresses the usage-based approach and argues against it and in favour of innate, domain-specific mechanisms of language. I agree with the latter, but notice that even if there is nothing unique to language in human cognition (so that we can generate and comprehend grammatical structures using domain-general and non-language-specific cognitive computations only), that would leave unaffected the claim that the uniquely human thought system is underwritten by the computational mechanisms that are revealed by studying whether subjects can recognise patterns based on particular phrase structure grammars on the Chomsky hierarchy. This is because there is nothing particularly linguistic about such formal mathematical hierarchies. The different levels on the hierarchy provide researchers with a measure of the complexity of structured patterns that a subject can and cannot recognise. They are a measure of cognitive complexity that can be manifested in any modality. This is the strength of using the Chomsky hierarchy, for it allows us to create structured patterns that provide a window into the underlying cognitive mechanisms in virtue of which such patterns can be generated and recognised. The work of Swaddle and Ruff (2004) and Stobbe et al. (2012), for example, used visual patterns generated by rules at different levels of computational complexity on the Chomsky hierarchy in order to better understand the underlying cognitive mechanisms of pattern perception in birds.

5. Concluding remarks

To sum up, the structural information and interpretations inherent in the strings generated by phrase structure grammars are indicative of a uniquely human kind of thought. Cognitive systems that cannot compute phrase structure grammars do not have access to such interpretations: they cannot think about the world in that way. In other words, the patterns (including the dependencies and structure therein) that a cognitive system can recognise embody a particular perspective on the world; the patterns the system is able to recognise reveals the way in which the system sees the world, the way in which it understands and interacts with the world. Thus, knowing the types of grammar on the Chomsky hierarchy (or, more generally, the types of rule-based patterns) that a subject can and cannot recognise reveals a great deal about the cognitive mechanisms of that subject. I argued that the subset of human thought that is not shared with animals can be accounted for by the underlying computational mechanisms of language (allowing for recursively-embedded, hierarchically-structured strings) that structure this subset of human thought in a particular way. This subset of thought is unavailable to cognitive systems that do not have access to the computations underlying natural language.

Acknowledgements

References

Abe, Kentaro, Watanabe, Dai, 2011. Songbirds possess the spontaneous ability to discriminate syntactic rules. Nat. Neurosci. 14, 1067–1074.
Adger, David, Svenonius, Peter, 2015. Linguistic explanation and domain specialization: a case study in bound variable anaphora. Front. Psychol. 6 (1421), 1–15.
Asoulin, Eran, 2013. The creative aspect of language use and the implications for linguistic science. Biolinguistics 7, 228–248.
Asoulin, Eran, 2016. Language as an instrument of thought. Glossa: J. Gen. Linguist. 1 (1), 1–23, 46.
Asoulin, Eran, 2019. Language And Scientific Explanation: Where does Semantics Fit in?. forthcoming Language Science Press, Berlin.
Bar-Hillel, Yehoshua, 1953. A quasi-arithmetical notation for syntactic description. Language 29, 47–58.
Berwick, Robert C., Okanoya, Kazuo, Beckers, Gabriel J.L., Bolhuis, Johan J., 2011a. Songs to syntax: the linguistics of birdsong. Trends Cognit. Sci. 15 (3), 113–121.
Berwick, Robert C., Pietroski, Paul, Yankama, Beracah, Chomsky, Noam, 2011b. Poverty of the stimulus revisited. Cogn. Sci. 35 (7), 1207–1242.
Bolhuis, Johan J., Okanoya, Kazuo, Scharff, Constance, 2010. Twitter evolution: converging mechanisms in birdsong and human speech. Nat. Rev. Neurosci. 11, 747–759.
Burton-Roberts, Noel, 2011. On the grounding of syntax and the role of phonology in human cognition. Lingua 121, 2089–2102.
Bybee, Joan, 2010. Language, Cognition, and Usage. Cambridge University Press, Cambridge.
Carstairs-McCarthy, Andrew, 1999. The Origins of Complex Language: An Inquiry into the Evolutionary Beginnings of Sentences, Syllables, and Truth. Oxford University Press, Oxford.
Carstairs-McCarthy, Andrew, 2000. The distinction between sentences and noun phrases: an impediment to language evolution? In: Knight, Chris, Studdert-Kennedy, Michael, Hurford, James R. (Eds.), The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form. Cambridge University Press, Cambridge, pp. 248–263.
Chomsky, Noam, 1956. Three models for the description of language. IEEE Trans. Inf. Theory 2 (3), 113–124.
Chomsky, Noam, 1957. Syntactic Structures. Mouton Publishers, The Hague.
Chomsky, Noam, 1959. On certain formal properties of grammars. Inf. Control 2, 137–167.
Chomsky, Noam, 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.
Chomsky, Noam, 1966. Cartesian Linguistics: A Chapter in the History of Rationalist Thought. Harper & Row, New York.
Chomsky, Noam, 1970. Remarks on nominalization. In: Jacobs, Roderick A., Rosenbaum, Peter S. (Eds.), Readings in English Transformational Grammar. Ginn, Waltham, MA, pp. 184–221.
Chomsky, Noam, 1981. Lectures on Government and Binding: The Pisa Lectures. Foris Publications Holland, Dordrecht.
Chomsky, Noam, 1995. The Minimalist Program. MIT Press, Cambridge, MA.
Chomsky, Noam, 2013. Problems of projection. Lingua 130, 33–49.
Chomsky, Noam, Miller, George A., 1958. Finite state languages. Inf. Control 1, 91–112.
Chomsky, Noam, Schützenberger, Marcel P., 1963. The algebraic theory of context-free languages. In: Paul, Braffort, Hirschberg, David (Eds.), Computer Programming and Formal Languages. North Holland Publishing Company, Amsterdam, pp. 118–161.
Christiansen, Morten H., Chater, Nick, 2015. The language faculty that wasn't: a usage-based account of natural language recursion. Front. Psychol. 6 (1182), 1–28.
Davidson, Donald, 1975. Thought and talk. In: Guttenplan, Samuel D. (Ed.), Mind and language: Wolfson College lectures 1974. Clarendon Press, Oxford, pp. 7–23.
Davidson, Donald, 1982. Rational animals. Dialectica 36 (4), 317–328.
de Villiers, Jill, 2014. What kind of concepts need language? Lang. Sci. 46, 100–114.
de Waal, Frans, 2016. Are We Smart Enough to Know How Smart Animals Are? W. W. Norton & Company, New York.
Dessalegn, Banchiamlack, Landau, Barbara, 2008. More than meets the eye: the role of language in binding and maintaining feature conjunctions. Psychol. Sci. 19 (2), 189–195.
Ding, Nai, Melloni, Lucia, Zhang, Hang, Tian, Xing, Poeppel, David, 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164.
Dummett, Michael, 1989. Language and communication. In: George, Alexander (Ed.), Reflections on Chomsky. Blackwell, Oxford, pp. 192–212.
Endress, Ansgar D., Cahill, Donal, Block, Stefanie, Watumull, Jeffrey, Hauser, Marc D., 2009. Evidence of an evolutionary precursor to human language affixation in a non-human primate. Biol. Lett. 5, 749–751.
Fitch, W. Tecumseh, Hauser, Marc D., 2004. Computational constraints on syntactic processing in a nonhuman primate. Science 303, 377–380.

Fitch, W. Tecumseh, Friederici, Angela D., 2012. Artificial grammar learning meets formal language theory: an overview. Phil. Trans. Biol. Sci. 367, 1933–1955.
Fitch, W. Tecumseh, Friederici, Angela D., Peter, Hagoort, 2012. Pattern perception and computational complexity: introduction to the special issue. Phil. Trans. Biol. Sci. 367, 1925–1932.
Fodor, Jerry A., 1975. The Language of Thought. Crowell, New York.
Fodor, Jerry A., 2003. Hume Variations. Clarendon Press, Oxford.
Fukui, Naoki, 2001. Phrase structure. In: Baltin, Mark, Collins, Chris (Eds.), The Handbook of Contemporary Syntactic Theory. Blackwell Publishers, Oxford, pp. 374–406.
Gazdar, Gerald, Klein, Ewan H., Pullum, Geoffrey K., Sag, Ivan A., 1985. Generalized Phrase Structure Grammar. Blackwell, Oxford.
Hilliard, Austin T., White, Stephanie A., 2009. Possible precursors of syntactic components in other species. In: Derek Bickerton & Eörs Szathmáry. MIT Press, Cambridge, MA, pp. 161–184. Biological Foundations and Origin of Syntax.
Hinzen, Wolfram, 2006. Internalism about truth. Mind Soc. 5, 139–166.
Hinzen, Wolfram, 2012. Human nature and grammar. Roy. Inst. Philos. Suppl. **70**, pp. 53–82.
Hinzen, Wolfram, 2013. Narrow syntax and the language of thought. Phil. Psychol. 26 (1), 1–23.
Hornstein, Norbert, 2009. A Theory of Syntax: Minimal Operations and Universal Grammar. Cambridge University Press, Cambridge.
Hornstein, Norbert, Pietroski, Paul M., 2009. Basic operations: minimal syntax-semantics. Catalan J. Ling. 8, 113–139.
Ludwig, Huber, Aust, Ulrike, Michelbach, G., Ölzant, S., Loidolt, Melanie, Nowotny, Regina, 1999. Limits of symmetry conceptualization in pigeons. Q. J. Exp. Psychol. B Comp. Physiol. Psychol. 52B (4), 351–379.
Jackendoff, Ray, 1977. X-bar syntax: a study of phrase structure. Linguistic Inquiry Monograph 2. MIT Press, Cambridge, MA.
Jackendoff, Ray, 1989. What is a concept, that a person may grasp it? Mind Lang. 4 (1), 68–102.
Jackendoff, Ray, 1990. Semantic Structures. MIT Press, Cambridge, MA.
Jackendoff, Ray, 2002. Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford.
Jackendoff, Ray, Wittenberg, Eva, 2014. What you can say without syntax: a hierarchy of grammatical complexity. In: Newmeyer, Frederick J., Preston, Laurel B. (Eds.), Measuring Grammatical Complexity. Oxford University Press, Oxford, pp. 65–82.
Jäger, Gerhard, Rogers, James, 2012. Formal language theory: refining the Chomsky hierarchy. Phil. Trans. Biol. Sci. 367, 1956–1970.
Joshi, Aravind K., 1985. Tree adjoining grammars: how much context-sensitivity is required to provide reasonable structural descriptions? In: Dowty, David R., Karttunen, Lauri, Zwicky, Arnold M. (Eds.), Natural Language Processing: Theoretical, Computational and Psychological Perspectives. Cambridge University Press, Cambridge, pp. 206–250.
Langacker, Ronald, 1987. Foundations of Cognitive Grammar, vol. 1. University of Chicago Press, Chicago.
Lasnik, Howard, Uriagereka, Juan, 2012. Structure. In: Ruth, Kempson, Fernando, Tim, Asher, Nicholas (Eds.), Handbook of the philosophy of science: Philosophy of linguistics, vol. 14. North Holland, Elsevier, Amsterdam, pp. 33–61.
Locurto, Charles, Fox, Maura, Mazzella, Andrea, 2015. Implicit learning in cotton-top tamarins (*Saguinus oedipus*) and pigeons (*Columba livia*). Learn. Behav. 43 (2), 129–142.
Malcolm, Norman, 1972. Thoughtless brutes. Proceedings and Addresses of the American Proceed. Addresses Am. Phil. Assoc. 46, 5–20.
McDowell, John, 1994. Mind and World. MIT Press, Cambridge, MA.
McGilvray, James, 2001. Chomsky on the creative aspect of language use and its implications for lexical semantics studies. In: Federica Busa, Bouillon, Pierrette (Eds.), The Language of Word Meaning. Cambridge University Press, Cambridge, pp. 5–27.
McGilvray, James, 2005. Meaning and creativity. In: McGilvray, James (Ed.), The Cambridge Companion to Chomsky. Cambridge University Press, Cambridge, pp. 204–222.
Pietroski, Paul M., 2008. Minimalist meaning, internalist interpretation. Biolinguistics 2 (4), 317–341.
Pietroski, Paul M., 2010. Concepts, meanings and truth: first nature, second nature and hard work. Mind Lang. 25 (3), 274–278.
Pietroski, Paul M., 2018. Conjoining Meanings: Semantics without Truth Values. Oxford University Press, Oxford.
Pollard, Carl, Sag, Ivan A., 1994. Head-Driven Phrase Structure Grammar. University of Chicago Press, Chicago.
Post, Emil, 1947. Recursive unsolvability of a problem of Thue. J. Symb. Log. 12, 1–11.
Pullum, Geoffrey K., Gazdar, Gerald, 1982. Natural languages and context-free languages. Linguist. Philos. 4, 471–504.
Rogers, James, Hauser, Marc D., 2010. The use of formal language theory in studies of artificial language learning: a proposal for distinguishing the differences between human and nonhuman animal learners. In: van der Hulst, Harry (Ed.), Recursion and Human Language. De Gruyter Mouton, Berlin, pp. 213–231.
Rosenbloom, Paul C., 1950. The Elements of Mathematical Logic. Dover, New York.
Ryle, Gilbert, 1968. A puzzling element in the notion of thinking. In: Strawson, Peter F. (Ed.), Studies In the Philosophy of Thought and Action. Oxford University Press, Oxford, pp. 7–23.
Saffran, Jenny, Hauser, Marc D., Seibel, Rebecca, Kapfhamer, Joshua, Tsao, Fritz, Cushman, Fiery, 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. Cognition 107, 479–500.
Slezak, Peter, 2002. Thinking about thinking: language, thought and introspection. Lang. Commun. 22, 353–373.
Stabler, Edward P., 2004. Varieties of crossing dependencies: structure dependence and mild context sensitivity. Cogn. Sci. 28, 699–720.
Stobbe, Nina, Westphal-Fitch, Gesche, Aust, Ulrike, Fitch, W. Tecumseh, 2012. Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). Phil. Trans. Biol. Sci. 367 (1598), 1995–2006.
Swaddle, John P., Ruff, Douglas A., 2004. Starlings have difficulty in detecting dot symmetry: implications for studying fluctuating asymmetry. Behaviour 141 (1), 29–40.
Tomasello, Michael, 2003. Constructing a Language: A Usage Based Theory of Language Acquisition. Harvard University Press, Cambridge, MA.
Tomasello, Michael, 2008. The Origins of Human Communication. MIT Press, Cambridge, MA.
Yip, Moira J., 2006. The search for phonology in other species. Trends Cognit. Sci. 10 (10), 442–446.