The Non-Hierarchical Nature of the Chomsky Hierarchy-Driven Artificial-Grammar Learning

Shiro Ojima & Kazuo Okanoya

Recent artificial-grammar learning (AGL) paradigms driven by the Chomsky hierarchy paved the way for direct comparisons between humans and animals in the learning of center embedding ([A[AB]B]). The A^nB^n grammars used by the first generation of such research lacked a crucial property of center embedding, where the pairs of elements are explicitly matched ([A1 [A2 B2] B1]). This type of indexing is implemented in the second-generation A^nB^n grammars. This paper reviews recent studies using such grammars. Against the premises of these studies, we argue that even those newer A^nB^n grammars cannot test the learning of syntactic hierarchy. These studies nonetheless provide detailed information about the conditions under which human adults can learn an A^nB^n grammar with indexing. This knowledge serves to interpret recent animal studies, which make surprising claims about animals' ability to handle center embedding.

Keywords: language evolution; animal cognition; syntactic hierarchy; artificial grammar; center embedding

1.   Center Embedding and A^nB^n Grammars

One of the properties that make humans unique among animals is language, which has several components including phonology, lexicon, and syntax. It has been debated how much of each of these components is shared between humans and non-human animals (Markman & Abelev 2004, Yip 2006). The component of syntax, which has been receiving much attention in the field of comparative cognition, instantiates linguistic knowledge describable in terms of a finite set of rules. That set of rules is called a grammar. Fitch & Hauser's (2004) seminal work tried to test which type of grammar non-human primates can learn. In doing so, they resorted to the distinction between a *finite-state grammar* and a *context-free grammar*, based on *the Chomsky hierarchy* (Chomsky 1957). Both these grammars can generate sets of surface strings such as "flying airplanes", but only the latter can generate phrase markers associated with surface strings, being able to differentiate between [VP flying [airplanes]] and [NP [flying] airplanes]. As in these

examples, natural-language sentences in the mind of a native speaker are hierar-
chically organized into units of phrases. The inadequacy of a finite-state grammar
as a model of human grammar can also be illustrated by sentences with *center
embedding* (e.g., The boy [the girl liked] smiled), which can be generated only by a
context-free grammar (or more powerful ones) (Chomsky 1957). The notion of
center embedding played a major role in the studies discussed below.

To compare humans and animals directly in a semantics-free fashion, Fitch
& Hauser (2004) expressed finite-state and context-free grammars as simple,
meaningless artificial grammars: a finite-state $(AB)^n$ grammar, which generated
sequences such as ABAB through local transitions (Figure 1a), and a context-free
A^nB^n grammar, which generated center-embedded, "hierarchical" structures such
as A[AB]B (Figure 1b). Because finite-state grammars had been observed in non-
human animals (Berwick *et al.* 2011, Fitch & Hauser 2004), the crucial question is
whether we can artificially induce, in animals, the learning of a "context-free"
A^nB^n grammar equipped with center embedding. This question bears direct rele-
vance to the evolutionary uniqueness of human language and generated a series
of artificial-grammar learning (AGL) studies driven by the Chomsky hierarchy.
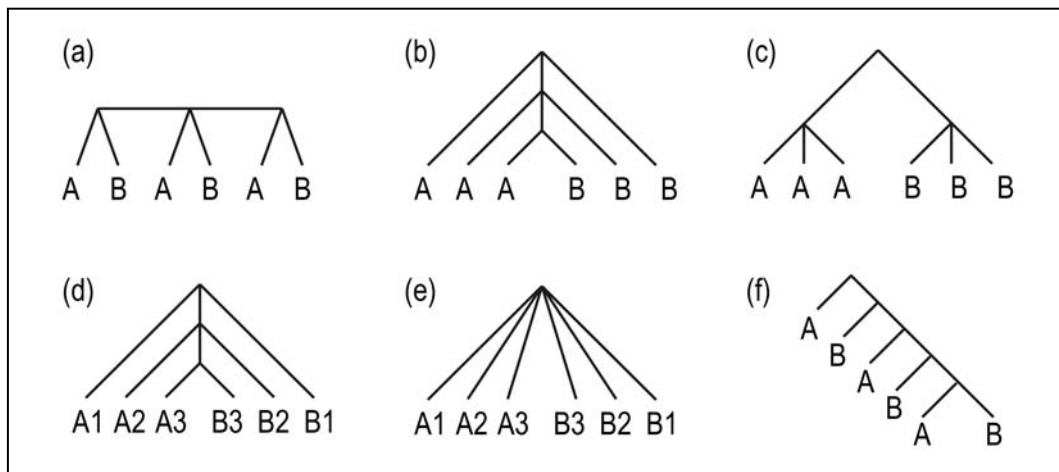

*Figure 1: Tree Diagrams Representing Artificial Grammars. (a, b) Original tree diagrams for (a)
the $(AB)^n$ "finite-state" grammar and (b) the A^nB^n "context-free" grammar (n = 3 here),
used in the first-generation AGL studies. (c) An alternative tree diagram for the first-
generation A^nB^n grammar. (d) A tree diagram for the indexed A^nB^n grammar used in the
second-generation studies. In this grammar, the pairs of As and Bs are matched from the
outer pairs inwards. Numbers (1, 2, 3) attached to As and Bs indicate which A is paired
with which B; for example, A1 is paired with B1, not with B2 or B3. (e) An alternative
representation of the indexed A^nB^n grammar, in which As and Bs are explicitly paired but
no hierarchical information is contained. As in (d), the numbers attached (1, 2, 3) show the
unique mapping relations between As and Bs, but nothing is hierarchically higher than
anything. (f) The "finite-state" $(AB)^n$ grammar represented as a tail-embedding,
hierarchical structure.*

The first generation of studies employing $(AB)^n$ and A^nB^n grammars tested
a variety of experimental subjects including humans (Bahlmann *et al.* 2006, Fitch
& Hauser 2004, Friederici *et al.* 2006), non-human primates (cotton-top tamarins)
(Fitch & Hauser 2004), and songbirds (European starlings and zebra finches)

(Gentner *et al*. 2006, van Heijningen *et al*. 2009), and reported striking evidence both for and against the human specificity of center embedding. Neuroimaging studies (Bahlmann *et al*. 2006, Friederici *et al*. 2006) claimed to have dissociated neural correlates of the processing of hierarchical structures (in an A^nB^n grammar) from those related to local transitions (in an $(AB)^n$ grammar).

However, "center-embedded" sequences such as AABB can be interpreted just as As followed by the same number of Bs (Corballis 2007a, 2007b, Perruchet & Rey 2005) (Figure 1c). A violation of this structure can be detected by simply counting the numbers of As and Bs (unequal numbers of As and Bs in an ungrammatical AABA string). Discrimination between "context-free" A^nB^n grammars and "finite-state" $(AB)^n$ grammars can be achieved in similar manners, for example, by counting the transitions between As and Bs (only one A-to-B transition in A^nB^n but multiple transitions in $(AB)^n$). Hence the task assigned to the subject in the first-generation A^nB^n studies could be performed independently of the way the string had been generated by the underlying grammar. The data reported in these studies do not count as evidence either for or against the human specificity of a context-free grammar. More recent, second-generation A^nB^n studies followed a proposal (Corballis 2007a, 2007b, Perruchet & Rey 2005) that the As and Bs in strings generated by an A^nB^n grammar be explicitly matched from the outside pairs inwards (not just [A[AB]B], but [A1[A2 B2]B1]). In the literature (see any of the second-generation A^nB^n studies introduced below), such a relationship is usually represented as in Figure 1d, where elements with the same number are intended to be paired (e.g., A*1* is paired with B*1*, not with B2 or B3). Center-embedded sentences in natural language show this type of pairwise dependencies. For example, native speakers of English would interpret "the boy the girl liked smiled" as having two subject-verb pairs, one (the girl-liked) embedded in the other (the boy-smiled). Hence this sentence not only is in the form of Subject Subject Verb Verb (SSVV) but also contains pairwise dependencies between subjects and verbs (S1 S2 V2 V1), in the minds of those who know the English syntax. An A^nB^n grammar explicitly indexed has been extensively used in the second-generation A^nB^n studies (Abe & Watanabe 2011, Bahlmann *et al*. 2008, de Vries *et al*. 2008, Fedor *et al*. 2012, Lai & Poletiek 2011, Mueller *et al*. 2010). Below we will call this new A^nB^n grammar with indexing an *indexed* A^nB^n grammar for short (but this should not be confused with the index of a context-free grammar (e.g., Salomaa 1969), which has been used in a totally different context).

Unlike the first-generation studies, these new experiments test whether the specific dependencies in the indexed AB pairs have been actually learned. After learning, the subject's sensitivity to grammatical strings having proper AB dependencies (e.g., A1 A2 A3 B3 B2 B1) and ungrammatical ones violating such dependencies (e.g., A1 A2 A3 B3 *B1 B2*) has been tested. Here the strategy of just counting the numbers of As and Bs does not help, because both grammatical and ungrammatical strings have the same number of As and Bs. The implementation of explicit indexing in the A^nB^n grammar has led many authors to assume that the second-generation studies have tested the learning and processing of *syntactic hierarchy* (Bahlmann *et al*. 2008, de Vries *et al*. 2008, Fedor *et al*. 2012, Fitch & Friederici 2012, Friederici *et al*. 2011, Lai & Poletiek 2011, Mueller *et al*. 2010).

2.   Hierarchy Is *Not* Involved

Despite the premises of these newer studies, syntactic hierarchy in a strict sense, we argue, has not been learned even in studies using indexed $A^n B^n$ grammars. It is true that an indexed $A^n B^n$ grammar introduces nested pairs and that participants are required to learn and process the dependencies between specific As and Bs. It is a different matter, however, whether humans interpret the strings generated by an indexed $A^n B^n$ grammar as containing syntactic hierarchy. Most of the second-generation $A^n B^n$ studies and a few review articles associate an indexed $A^n B^n$ grammar with the hierarchical structure building of natural language (Abe & Watanabe 2011, Bahlmann *et al*. 2008, de Vries *et al*. 2008, Fitch & Friederici 2012, Friederici *et al*. 2011, Lai & Poletiek 2011, Mueller *et al*. 2010). These papers graph-ically represent the indexed $A^n B^n$ grammar as in Figure 1d. This representation is misleading, in that it gives us an impression that the outer pairs are hierarchically higher than the inner pairs, but such information is not provided during learning as part of familiarization strings and thus cannot be learned. A more accurate re-presentation of the second-generation $A^n B^n$ grammar is in Figure 1e, where infor-mation about pairs is present but information about hierarchy is not. Here, ele-ments with the same number are in a pair (e.g., A*1* is paired with B*1*, not with B2 or B3) as in Figure 1d, but no hierarchy is contained. As long as no hierarchical information is conveyed, the learning of an indexed $A^n B^n$ grammar in the second-generation studies is the learning of center-embedded or nested pairs, but not the learning of hierarchy.

More generally, we cannot make distinctions in hierarchy between "finite-state" $(AB)^n$ grammars and "context-free" $A^n B^n$ grammars, based solely on famili-arization strings. The English sentence "Bob believes Mary came" can be described as noun verb noun verb, or ABAB if A is a noun and B is a verb, but is fully hierarchical in the mind of a native English speaker, who will interpret this sentence as consisting of a higher main clause and a lower embedded clause, as in [Bob believes [Mary came]]. In terms of hierarchy, $(AB)^n$ strings and $A^n B^n$ strings in the studies reviewed here are no different; neither have inherent hierarchical structure and can thus be interpreted either as flat or as hierarchical, depending on lexical items which one imagines inserting. If an AAABBB string is interpreted as having a center-embedded, hierarchical [A[A[AB]B]B] structure, then an ABABAB string can also be interpreted as having a tail-embedded, hierarchical [AB[AB[AB]]] structure (Figure 1f).

In our view, both the first- and the second-generation studies have made the same mistake. The artificial grammar which generated a string is equated with the psychological process involved in the processing of that string, but these two are not the same (Lobina 2011). It is certainly true that hierarchy has been necessary to describe the knowledge of language (language competence); concepts such as *c-command* (Figure 2a) based on syntactic hierarchy have been indispensable for the accounts of many grammatical constructions (Carnie 2006, Chomsky 1981). However, there are examples of non-hierarchical, flat sequences in natural language. Here we take *negation* as an example. Syntactic hierarchy can be easily seen in constructions involving negation. In the sentence "Never dis-agree!", the first negator "never" is hierarchically higher than the second negator

"dis–", which is contained in the word "disagree" (Figure 2b). This double negation leads to a (hesitated) affirmative interpretation of the verb "agree" (negative of negative → affirmative). However, we cannot always assume hierarchy of this sort for each negation. For example, it is wrong to do so for the sentence "Never, never say that!". If this sentence had a fully hierarchical representation as in Figure 2c, the first "never" would erroneously negate the rest of the sentence involving the second "never" and would thus lead to an affirmative interpretation of "say that", which is obviously the wrong interpretation. The sentence is singly, not doubly, negative in meaning, and we should postulate a flat representation for the part "never, never" as in Figure 2d.
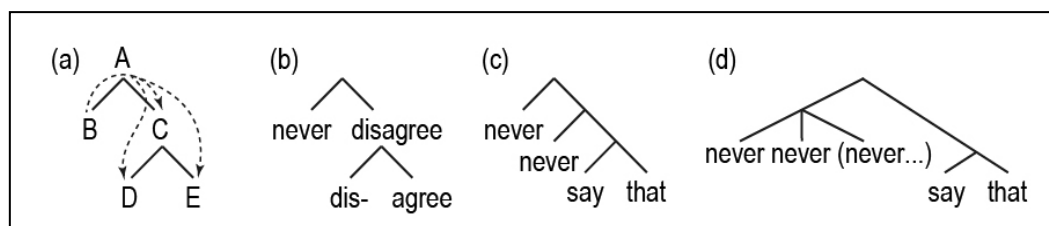

Figure 2: Tree Diagrams Representing Some Grammatical Relations. (a) C-command. B and C c-command each other, while B asymmetrically c-commands D and E (which do not c-command B). (b-d) Interpretation of negation depends on syntactic hierarchy. Double negation (negative of negative) leads to a (hesitated) affirmative meaning, as in (b). However, we cannot always assume hierarchy between two negators; the representation in (c) cannot be correct. We should give "never, never" a non-hierarchical, flat representation as in (d).

For some other constructions, even theoretical linguists did not know for sure (and thus had to debate) whether hierarchy should be assumed. Japanese, which has relatively flexible word order, had once been thought to have non-hierarchical, flat structure in a clause (Hale, 1980, 1982) (Figure 3a).
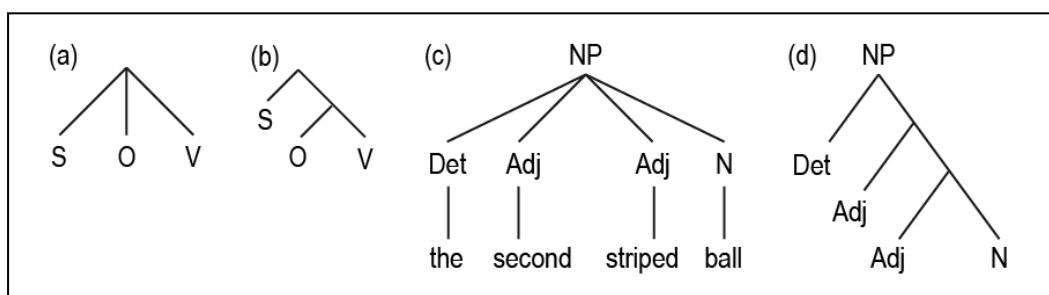

Figure 3: Examples of natural-language constructions, for which theoretical linguists were once divided between hierarchical and non-hierarchical representations. (a) Proposed flat, non-hierarchical representation for subject (S) – object (O) – verb (V) sentences in Japanese. (b) Hierarchical representation for Japanese SOV sentences. (c) Flat noun phrase (NP) proposed for child English speakers. (d) Hierarchical noun phrase.

Later research denied this view and showed that Japanese was as hierarchical as English (Saito & Hoji 1983) (Figure 3b). Also, English-speaking

children were once thought to have a non-hierarchical, flat noun phrase (NP) as in Figure 3c. Only a more careful analysis of children's language comprehension revealed that their noun phrase was hierarchical like adults' (Figure 3d) (Crain & Lillo-Martin 1999). English-speaking children form complex interrogatives involving a relative clause in a structure-preserving way as predicted by theoretical accounts of English phrase structure, but this was revealed only through clever experiments (Crain & Nakayama 1987). By analogy, it should not be taken for granted that the subject's mental representations of artificial-grammar sentences are hierarchical; it needs to be demonstrated.

Even if this can be achieved, the actual use of such knowledge required by specific task demands may not depend on the processing of hierarchy. A recent hypothesis questions the involvement of hierarchy in real-time use of natural language, even if its mental representation may still be hierarchical (Frank *et al*. 2012). According to this hypothesis, the involvement of hierarchy must be shown at both the level of mental representations (competence or language knowledge) and the level of real-time processing (performance or knowledge use). The competence/performance distinction is one of the most fundamental concepts in generative linguistics (Chomsky 1965), which is almost exclusively concerned with competence, or the speaker/hearer's internal representation of finite rule sets that generate sentences. Keeping this distinction in mind is not just useful but sometimes necessary, especially where linguists and non-linguists discuss things on a common ground, a primary example of which is AGL studies. The importance of this distinction in experimental studies has been recently reiterated elsewhere (Petersson & Hagoort 2012). In A^nB^n studies, evidence for the involvement of hierarchy in the learning of an A^nB^n grammar has not been provided either at the level of performance or at the level of competence. After all, it has not been studied how the subject processes input strings internally, and we simply cannot know the nature of the internal representations used by the subject in the processing of those strings.

Perhaps those who claim to have studied syntactic hierarchy by using the indexed A^nB^n grammar assume that this grammar automatically introduces the types of hierarchy shown by natural-language sentences conforming to the general pattern of A^nB^n. There are many such sentences, and we can easily see what kind of hierarchy is present in each of them. A typical example would be "John, who Mary liked, smiled.", whose (simplified) tree diagram is shown in Figure 4a. Here the inner sentence "who Mary liked" is attached to the left (to the side of "John"), giving additional information about the subject "John". In a similar sentence, "John, when Mary came, smiled.", that is not the case. As shown in Figure 4b, the inner sentence "when Mary came" is attached to the right (to the side of "smiled") and does not modify the subject "John". If we add another sentence "Bill did *so* too" at the end, it will mean "Bill *smiled when Mary came*, too". This suggests that "when Mary came" is tied to the verb "smiled". One thing we can say here is that in certain natural-language A^2B^2 (more generally A^nB^n) sentences, the inner pair, A2 and B2, is attached either to the left (Figure 4c) or to the right (Figure 4d), but not to the center (Figure 4e) or to nowhere (Figure 4f). The *center*-embedded part is only superficially in the center and is actually attached to just one side, and bears no direct relation to the other.

Even if the direction of attachment is the same, how attachment is done may differ among natural-language sentences. Another typical example of the A^nB^n pattern is found in sentences such as "If either S or S, then S", where S is for Sentence (Chomsky 1957). This sentence has syntactic hierarchy (Figure 4g) that is similar to the one of "John, who Mary liked, smiled". In both these sentences, the inner sentence is attached to the left (to the side of "John" or "if"). However, in the "if" sentence, the inner sentence "either S or S" cannot be deleted ("If then …" is ungrammatical), while in the "John" sentence, the inner sentence "who Mary liked" can be deleted ("John smiled" is grammatical). Here we have the distinction between *complements* and *adjuncts*. "If" must have a complement to stand alone as a syntactic unit and thus requires a sentence. "John" itself is a proper syntactic unit, and we can adjoin something to it but do not have to. Syntactically, complements and adjuncts are in different hierarchical positions (Figure 4h) and are known to behave differently (see Radford 1988 for examples).
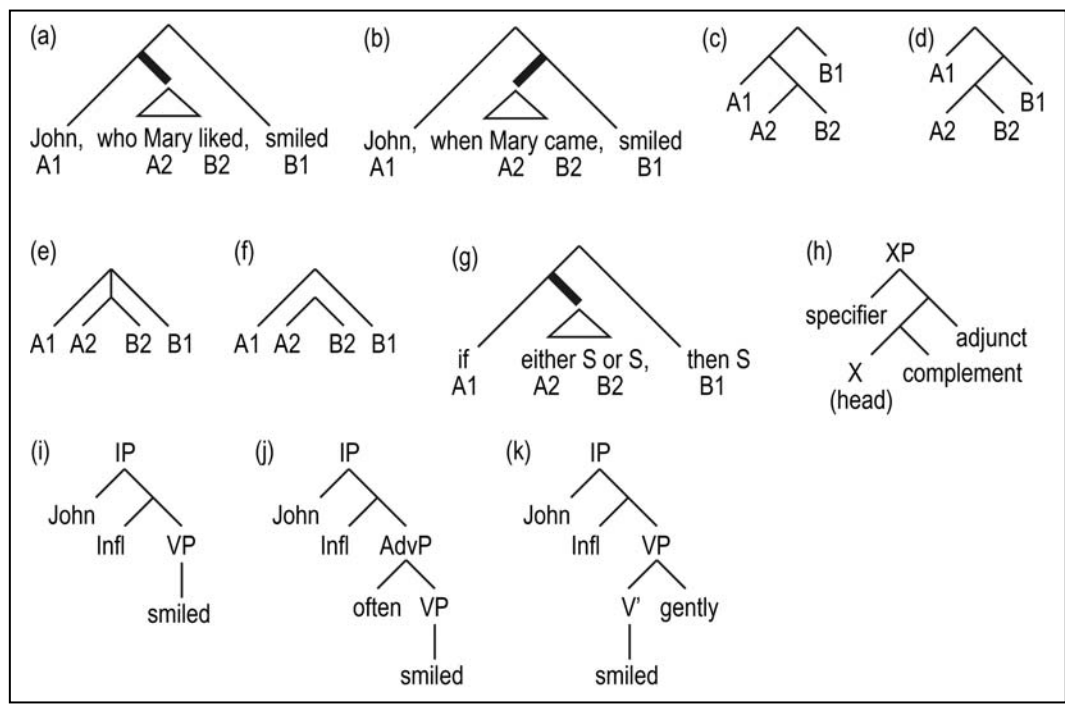

Figure 4:   Compatibility between strings and hierarchical representations. In natural-language sentences that conform to the A^nB^n pattern, what is inserted between A and B may actually be attached to the left as in (a) or to the right as in (b). In natural language, the inner AB is attached to the outer A (c) or to the outer B (d), but is not hanging from the center (e) or hanging from nowhere (f). An English sentence containing pairs of "if-then" and "either-or" (g) looks similar in hierarchy to the (a) sentence containing a relative clause, but in (g), the part containing "either-or" is a complement whereas in (a), the relative clause is an adjunct. (h) Complements and adjuncts occupy different hierarchical positions in a phrase. (i) Natural-language sentences conforming to the simple AB pattern may still have hierarchy. Inserting something between A and B (ACB) may create more hierarchy (j), but adding something after B (ABC) may do so, too (k).

It should be clear that the A^nB^n pattern in artificial grammars is compatible with many kinds of hierarchical representations found in natural-language

sentences. Just by inserting a pair of AB inside another does not specifically select one of these. In fact, whether and what syntactic hierarchy is created by doing so cannot be known. One may be pleased that at least some hierarchy is created, but just to have syntactic hierarchy of some sort, we do not have to have nested pairs. Just a simple AB pair may be hierarchical in natural language; hierarchy is present even in "John smiled" (Figure 4i). If we put something between A and B ("John *often* smiled"), the sentence may (or may not) have more hierarchy (Figure 4j), but if we put something after the AB pair ("John smiled *gently*"), we may achieve the same thing (Figure 4k). Hence the nesting of AB pairs in an artificial grammar is not special in its compatibility with hierarchical representations of natural-language sentences. Strings without nesting are also compatible with hierarchy present in natural-language sentences.

Some previous AGL research has tested AXB grammars. If, as assumed in the A^nB^n studies, putting a pair of elements (A2 B2) between the two elements of another pair (A1 B1) automatically introduces syntactic hierarchy, then we should equally assume that syntactic hierarchy is present in (and thus can be studied by) a string such as A1 X B1, where X can vary freely while A and B are in a non-adjacent dependency. The A1 A2 B2 B1 string is a special case of this, if the inner pair (A2 B2) is regarded as a unit (X). Artificial AXB grammars have been frequently used to study the learning of non-adjacent dependencies (Newport & Aslin 2004, Newport *et al*. 2004), but have never been claimed to tap syntactic hierarchy. Obviously, X in AXB is not hierarchically lower than A and B, in the absence of explicit evidence that it is. Likewise, the inner pair A2 B2 in A1 A2 B2 B1 is not hierarchically lower than the outer pair A1 B1. In effect, syntactic hierarchy has not been studied in either the first- or the second-generation A^nB^n studies. Some also argue that the Chomsky hierarchy on which the A^nB^n studies are based is not relevant to the neurobiological studies of language at all (Petersson *et al*. 2012).

To sum up, there is no strong evidence that syntactic hierarchy is involved in the learning and processing of either the first-generation un-indexed A^nB^n grammars or the second-generation indexed A^nB^n grammars.

3.   The Learnability of Indexed A^nB^n Grammars

The second-generation A^nB^n studies mainly addressed the issue of under what conditions human adults can learn an indexed A^nB^n grammar. The learnability of this grammar revealed recently will inform comparisons between humans and animals in center-embedding learning, but without reference to syntactic hierarchy. The original study which first introduced the second-generation indexed A^nB^n grammar (Perruchet & Rey 2005) reported that the dependencies implemented were impossible even for human adults to learn, if the learning procedure was the same as in the original A^nB^n study (Fitch & Hauser 2004). Inspired by this finding, most of the second-generation A^nB^n studies (Bahlmann *et al*. 2008, de Vries *et al*. 2008, Fedor *et al*. 2012, Lai & Poletiek 2011, Mueller *et al*. 2010) addressed the issue of the learnability of indexed A^nB^n grammars in human adults, and have now managed to describe under which conditions humans

succeed or fail in A^nB^n learning. Table 1 (next page) summarizes the key characteristics of 18 experiments which employed an indexed A^nB^n grammar. The results of A^nB^n learning in these experiments are of three types: failure, success, and success possibly aided by a task-taking strategy such as "repetition detection" (de Vries *et al.* 2008). The discussions below exclude the cases where a strategy might have been used.

3.1. Explicit A^nB^n learning in the visual modality

As of now, most is known about the *explicit* (as opposed to *implicit*) learning of an indexed A^nB^n grammar in the *visual* (as opposed to *auditory*) modality (Bahlmann *et al.* 2008, de Vries *et al.* 2008, Fedor *et al.* 2012, Lai & Poletiek 2011). The conditions under which human adults' explicit A^nB^n learning in the visual modality tends to be successful include the following: (1) the subject actively searches for rules during familiarization, (2) negative feedback is given about the correctness of the rules the subject found, (3) familiarization strings contain "0-LoE" items and are presented in a "staged" manner, (4) inherent phonological or semantic cues exist between the dependent elements of As and Bs, (5) the level of embedding is one ([A[AB]B]) or two ([A[A[AB]B]B]) (but not three or more), and (6) learning continues for at least 20–30 minutes, and the subject is given 200–300 sentences. Each of these conditions will be discussed in more detail below.

In successful explicit A^nB^n learning in the visual modality, the subject *actively searched for rules* during familiarization phases. Typically, the subject was told that familiarization strings (all grammatical) had been generated by rules, and while those strings were being presented, the subject tried to find those underlying rules.

Negative feedback is provided during rule-testing phases, which are part of learning. During rule-testing phases, the subject can test the correctness of the rules they found during familiarization. Both grammatical and ungrammatical strings are presented, and the subject has to judge each of them for grammaticality. Based on feedback on each judgment, the subject has chances to modify their own rules.

Zero-LoE items (0 level of embedding items) are strings that do not have embedding, that is, simple AB strings (Lai & Poletiek 2011). Zero-LoE items help the subject quickly find out which A is paired with which B. However, for this knowledge to be effective in the induction of the embedding structure of 1-LoE and 2-LoE items, 0-LoE items must be learned first, that is, before 1-LoE and 2-LoE items (Lai & Poletiek 2011). Input that is presented according to the level of embedding (0-LoE → 1-LoE → 2-LoE) is called *staged*. In A^nB^n learning, staged input greatly helps the subject induce the internal structure of complex strings. However, for facilitation to occur, 0-LoE items and staged input must be

Study & Condition	Result	N of A^nB^n	Active rule search	Feed-back	Staged input	0-LoE items	Cues	Exposure length
VISUAL								
Bahlmann 2008	Success*	3	√	√	√	√	√	35 min
de Vries 2008								
Exp. 1, Hier-Scram	Failure	4	√	√				50 min
Exp. 2, Hier-Scram+Rep	Success*	3	√	√			√	50 min
Exp. 2, Hier-Scram	Failure	3	√	√			√	50 min
Lai 2011								
Exp. 1, SS	Success	3	√	√	√	√	√	30 min, 12 blocks
Exp. 1, random	Failure	3	√	√		√	√	30 min, 12 blocks
Exp. 2, SS	Failure	3	√	√	√		√	8 blocks (20 min#)
Exp. 2, random	Failure	3	√	√			√	8 blocks (20 min#)
Fedor 2012								
WS	Success	3	√	√	√	√	√	7.28 blocks
WR	Success	3	√	√	√	√		12.27 blocks
NR1	Success	3	√	√	√	√		16.94 blocks
NR2	Success	3	√	√	√	√		20.25 blocks
Udden 2012, Exp. 2	Success	3						9 sessions (each < 30 min)
AUDITORY								
Perruchet 2005	Failure	3						3 min
Mueller 2010								
C1, no boundary	Failure	2	√				√	12 min
C2, prosody	Success	2	√				√	13 min#
C3, prosody + pause	Success	2	√				√	18 min#
C4, prosody + pause	Success	2	√				√	21 min

Table 1. Characteristics of center-embedding learning in the second-generation A^nB^n studies. Success here is defined as "at least above chance". Success (with an asterisk) means success possibly aided by a strategy. One "block" in column "exposure length" consists of about 10 familiarization strings and about 10 test strings. Abbreviations: 0-LoE = zero level of embedding. Exp. = experiment. Hier-viol = hierarchical-violation. Scram-viol = scrambled-violation. Hier-Viol + Rep = hierarchical-violation + repetition. SS = starting small. WS = words with semantic association paired. WR = words randomly paired. NR1 & 2 = non-words randomly paired. C1 – 4 = condition 1 - 4. # Our estimates.*

combined. If 0-LoE items are presented together with 1- and 2-LoE items from the beginning, facilitation does not occur. Similarly, if input is staged but 0-LoE items are not used (just 1-LoE → 2-LoE), facilitation does not occur, either (Lai & Poletiek 2011). In natural language, input that is staged according to complexity is considered to facilitate the learning of complex structures (Elman 1993). *Starting small*, in the form of staged input or others, may be a natural property of children's first language acquisition (Newport 1990). A theoretical account (Poletiek 2011, Poletiek & Lai 2012) considers the effect of staged input in terms of how much grammatical information is contained in the input strings.

Inherent cues about pairings have been shown to facilitate A^nB^n learning. In many of the second-generation studies, *phonological* cues are provided as to the pairings of elements (Bahlmann *et al*. 2008. de Vries *et al*. 2008, Lai & Poletiek 2011, Mueller *et al*. 2010). An example string would be "de gi ko tu", where "de" and "tu" are paired (outer pair), and "gi" and "ko" are paired (inner pair). The two elements in each pair agree in a phonological feature such as place of articulation (/d/ & /t/, /g/ & /k/). Semantic cues (e.g., semantically related real words such as "you" and "me" paired) greatly facilitate A^nB^n learning (Fedor *et al*. 2012). Facilitation also occurs, to a lesser extent, when real words are randomly paired (e.g., "me" and "lake" for A and B). In the absence of any useful cues, learning occurred to some extent. It is notable, however, that under this condition, 25% of the subjects (normal adults) could not learn pairings, given as many as 400 training sentences (Fedor et al., 2012). Hence to ensure 100% success, some kind of inherent cues about pairings seem to be necessary.

The learning of an A^3B^3 (2-LoE) grammar has been demonstrated, but there is no report on the learning of A^4B^4 (3-LoE), which had been studied in the first-generation studies (Bahlmann *et al*. 2006, Friederici *et al*. 2006). These tendencies may correspond to the limitations on multiple uses of embedding observed in natural language corpora (Karlsson 2007).

Learning continued for at least 20–30 minutes, and 200–300 sentences were presented to the subject. In the earliest A^nB^n studies (Fitch & Hauser 2004, Perruchet & Rey 2005), exposure to the grammar was as short as a few minutes. The learning of an indexed A^nB^n grammar may not be possible in such a short time, even if the other conditions are met.

3.2. Implicit A^nB^n learning in the visual modality

At least one study (Udden *et al*. 2012) reports that an indexed A^nB^n grammar presented in the visual modality can be *implicitly* learned. In this experiment (Experiment 2 in the article), most factors that have been reported to facilitate explicit A^nB^n learning in the visual modality are not used. The subject was not engaged in active rule search and was not given negative feedback as to the correctness of their grammaticality judgments (judgments were not done as part of learning). Zero-LoE items were not provided, and input was not presented in a staged manner. Inherent phonological or semantic cues were not present for the AB dependencies. Despite these seemingly disadvantageous features, effects of learning were observed. The secret may lie in the length of learning. The subject went through nine sessions in a period of two weeks. During one session (max.

30 minutes), the subject was shown 100 grammatical strings, which they had to type using a keyboard. In total, 900 strings were presented. This is several times as many as the number of familiarization strings used in the explicit-learning studies. Hence, the implicit learning of an indexed A^nB^n grammar in the visual modality seems to be possible in human adults, given a far larger number of familiarization strings than in explicit learning, even if the facilitative factors already known are not used. We should also note that only this study (Udden *et al.* 2012) used the whole-sentence presentation, where the subject could see the entire sentence on the display, as opposed to successive presentation, employed by the other visual studies, where the sentence was presented in an element-by-element manner.

3.3. A^nB^n learning in the auditory modality

The second-generation studies conducted in the auditory modality are a minority, and it is difficult to make a generalization. There may be special effects of sensory modalities (i.e., visual vs. auditory), but this needs to be confirmed by future research. In the auditory modality, only the learning of 1-LoE (i.e., A^2B^2) has been shown (Mueller *et al.* 2010), although in the visual studies, the learning of 2-LoE (i.e., A^3B^3) is reported to be possible (Fedor *et al.* 2012, Lai & Poletiek 2011). This may reflect general difficulty with comprehending embedding in speech streams (Karlsson 2007). Alternatively, methodological differences may be at issue here. In successful A^2B^2 learning in the auditory modality (Mueller *et al.* 2010), input was not staged, and 0-LoE items were not presented. Negative feedback was not used, either, although the subject actively searched for rules in the input. Above-chance learning occurred in conditions where the boundaries of strings (sentences) are marked by prosody or by both prosody and pauses. The artificial grammar in this study utilized phonological cues about pairings. In the auditory modality, center-embedding learning without such cues has not been demonstrated.

As we saw above, the learning of an indexed A^nB^n grammar is possible only under highly specific conditions, even in human adults. When one or more of those conditions are not met, learning becomes difficult or impossible. The findings of the second-generation A^nB^n studies on humans constitute a baseline against which the behavior of non-human animals should be judged.

3.4. Songbirds

Currently there are few animal studies on the learning of center embedding in the framework of the second-generation, indexed A^nB^n grammar. To make reliable comparisons between humans and animals, we simply need more research on animals. As we have seen above, much research has already been conducted on humans, and much knowledge about the learnability of an indexed A^nB^n grammar in humans has accumulated. Future research should build on such human research and test animals. That said, we now turn to the few exceptional animal studies that have been published recently.

A songbird species (Bengalese finch) has been claimed to have learned an

indexed A^nB^n grammar implicitly and spontaneously (without training or reinforcement), to the level of A^3B^3, via completely passive exposure (Abe & Watanabe 2011). Birds were not trained on ungrammatical strings and were not given positive or negative feedback. Familiarization strings were not given in a staged-input manner. No inherent cues were present in the AB dependencies. Birds were familiarized to grammatical strings only during one session of 60 minutes. In this type of short-exposure paradigm, the learning of A^3B^3 via passive exposure, without negative feedback, without staged input, without inherent cues, has not been demonstrated even in human adults. In fact, humans' learning of A^3B^3 in a meaningless artificial grammar in the auditory modality has not been shown with any learning procedure. Only in a long-exposure paradigm, involving nine sessions of exposure spreading over a period of two weeks, have humans been shown to learn a visual A^3B^3 grammar implicitly (Udden *et al*. 2012).

The claim made in first-generation research that songbirds (starlings) can learn to discriminate grammars with or without center embedding (Gentner *et al*. 2006) merely meant that songbirds, after intensive training, could do something that humans could easily do (without any training, in this particular case). If Bengalese finches can really learn A^3B^3 implicitly and spontaneously in such a short time, this finding can be interpreted as having gone a step further; without any training, birds can do something that humans cannot, or at least have not been proven to be capable of. A close inspection of the test strings used in the Bengalese finch study (Abe & Watanabe 2011) suggested the possibility that the finches behaved according to acoustic similarity among stimuli, rather than grammar (Beckers *et al*. 2012). Methodologically more rigorous research is necessary to precisely describe Bengalese finches' ability to learn center embedding (ten Cate & Okanoya 2012).

3.5. Non-human primates

A recent study reports that non-human primates have a spontaneous tendency to produce center embedding (Rey *et al*. 2102). In contrast to all the other studies above, the subjects in this study, baboons, were not exposed to center-embedding strings at all, and hence did not learn center embedding from external input. They learned pairs of meaningless visual shapes displayed on the monitor. The shapes appeared at random locations, and the baboons were conditioned to touch the correct combinations of shapes in the correct orders (e.g., touch A1 then touch B1). During training, they were required to sequentially touch two shapes at a trial. During test sessions, they were prompted, for the first time, to touch four shapes. For testing, they were shown, for example, A1 first, A2 second, and later, B1 and B2 simultaneously. In this case, they had to touch A1 first and A2 second, but for the latter part they had choices as to which of the stimuli to touch in what order. Specifically, they could choose to touch B2 and then B1 (A1→A2→B2→B1, consistent with center embedding) or B1 and then B2 (A1→A2→B1→B2, not consistent with center embedding). Results show that baboons have a spontaneous tendency to produce more responses which are consistent with center embedding (A1→A2→B2→B1), than those which are not (A1→A2→B1→B2).

This study is special in the second-generation A^nB^n studies, in the sense that it tested whether responses consistent with center embedding are *produced* spontaneously, without conditioning. One might argue that non-human primates' preference to produce center embedding is the evolutionary origin of humans' center embedding, but as of now, no data are available on whether humans show the same preference when put in the same situation. Previous studies on humans have not looked at the issue of center embedding from this perspective. Moreover, it is possible that the preference to put visual shapes in an order consistent with center embedding is not related to center embedding seen in the human grammar at all. In the human grammar, center embedding appears through the interaction of the so-called *head directions* of phrases. Different languages may have different directions for heads (Chomsky 1981, 1986). For example, English is a head-initial language and Japanese is a head-final language. A center-embedding sentence in English like "The boy the girl liked smiled", if directly translated, will not have a center-embedding structure in Japanese (the words will be ordered as in "the girl liked (whom) the boy smiled", to produce the same relative-clause structure with the same meaning). Hence it is one's grammar that determines whether center embedding must be used or cannot be used. Whether the appearance of center embedding in human language has its evolutionary origin in the reported preference of non-human primates to produce center embedding should be supported by further research.

3.6. Comparisons between humans and animals

As we saw above, the second-generation A^nB^n studies on humans as a whole show that the learning of an indexed A^nB^n grammar is very difficult even for human adults and is possible only under specific conditions. It is particularly important to note that this learning is difficult even if humans are *required* to do it. However, both two studies on non-human animals (songbirds and non-human primates) we have just discussed above are in favor of the view that animals also have an ability to handle center embedding. Both those studies provide evidence for this view from animals' spontaneous behavior, without using conditioning or reinforcement. On the one hand, humans' learning of center embedding is difficult even if required. On the other, animals are claimed to have demonstrated center embedding even without being required. This would make more sense if it was exactly the other way around. We have to say that we are in a somewhat odd situation where non-human animals without natural language are claimed to be able to handle a linguistic operation that is difficult even for humans who have natural language. As things currently stand, we are yet to see convergence between animal studies and human studies on the issue of the learnability of an indexed A^nB^n grammar. To move the field forward, each of the two lines of studies should respect the methodological details of the other. Fine methodological details can influence the outcome of AGL (Pena *et al*. 2002). Although the second-generation A^nB^n studies on humans and animals reviewed here all implemented dependencies between As and Bs in a broad sense, humans and animals have not been compared using exactly the same methodologies. Only carefully designed studies can resolve the discrepancy that currently exists

between the results of human studies and those of animal studies. The shortage of evidence on animals is also a notable feature of the second-generation studies. Many more studies on animals will be appreciated.

4.   Conclusion

The recent, second-generation A^nB^n studies have tested for dependencies in As and Bs, which the older, first-generation A^nB^n studies had not. This led to a currently standard view that the second-generation indexed A^nB^n grammars can be used to test syntactic hierarchy. We argue against this view and claim that syntactic hierarchy cannot be tested with the current experimental setups employed by the A^nB^n studies. These studies offer opportunities to compare humans and animals, within this limitation. The second-generation studies show that the learning of an indexed A^nB^n grammar is fairly difficult even for human adults and is possible under highly specific conditions. This observation is difficult to reconcile with the recent claims that center embedding is observed in non-human animals' spontaneous behavior. Carefully designed comparisons between humans and animals are awaited.

References

Abe, Kentaro & Dai Watanabe. 2011. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience* 14(8), 1067–1074.

Bahlmann, Jörg, Thomas C. Gunter & Angela D. Friederici. 2006. Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience* 18(11), 1829–1842.

Bahlmann, Jörg, Richard I. Schubotz & Angela D. Friederici. 2008. Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42(2), 525–534.

Beckers, Gabriël J. L., Johan J. Bolhuis, Kazuo Okanoya & Robert C. Berwick. 2012. Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *Neuroreport* 23(3), 139–145.

Berwick, Robert C., Gabriël J. Beckers, Kazuo Okanoya & Johan J. Bolhuis. 2012. A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience* 4, 5.

Berwick, Robert C., Kazuo Okanoya, Gabriel J. L. Beckers & Johan J. Bolhuis. 2011. Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences* 15(3), 113–121.

Carnie, Andrew. 2006. *Syntax: A Generative Introduction*. Oxford: Blackwell Publishers.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Corballis, Michael C. 2007a. On phrase structure and brain responses: A comment on Bahlmann, Gunter, and Friederici (2006). *Journal of Cognitive Neuroscience* 19(10), 1581–1583.

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

Crain, Stephen & Diane Lillo-Martin. 1999. *An Introduction to Linguistic Theory and Language Acquisition*. Malden, MA: Blackwell Publishers.

Crain, Stephen & Mineharu Nakayama. 1987. Structure dependence in grammar formation. *Language* 63, 522–543.

de Vries, Meinou H., Padraic Monaghan, Stefan Knecht & Pienie Zwitserlood. 2008. Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition* 107(2), 763–774.

Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48(1), 71–99.

Fedor, Anna, Máté Varga & Eörs Szathmáry. 2012. Semantics boosts syntax in artificial grammar learning tasks with recursion. *Journal of Experimental Psychology-Learning Memory and Cognition* 38(3), 776–782.

Fitch, W. Tecumseh & Angela D. Friederici. 2012. Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B–Biological Sciences* 367(1598), 1933–1955.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656), 377–380.

Frank, Stefan L., Rens Bod & Morten H. Christiansen. 2012. How hierarchical is language use? *Proceedings of the Royal Society B* 279, 4522–4531.

Friederici, Angela D., Jörg Bahlmann, Roland Friedrich & Michiru Makuuchi. 2011. The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* 5(1–2), 87–104.

Friederici, Angela D., Jörg Bahlmann, Stefan Heim, Ricarda I. Schubotz & Alfred Anwander. 2006. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America* 103(7), 2458–2463.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440(7088), 1204–1207.

Hale, Ken. 1980. Remarks on Japanese phrase structure: Comments on the papers on Japanese syntax. *MIT Working Papers in Linguistics* 2, 185–203.

Hale, Ken. 1982. Preliminary remarks on configurationality. *Proceedings of NELS [North Eastern Linguistic Society]* 12, 86–96.

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43(2), 365–392.

Lai, Jun & Fenna H. Poletiek. 2011. The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition* 118(2), 265–273.

Lobina, David J. 2011. Recursion and the competence/performance distinction in

AGL tasks. *Language and Cognitive Processes* 26(10), 1563–1586.

Markman, Ellen M. & Maxim Abelev. 2004. Word learning in dogs? *Trends in Cognitive Sciences* 8(11), 479–481.

Mueller, Jutta L., Jörg Bahlmann & Angela D. Friederici. 2010. Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science* 34(2), 338–349.

Newport, Elissa L. 1990. Maturational constraints on language learning. *Cognitive Science* 14, 11–28.

Newport, Elissa L. & Richard N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48(2), 127–162.

Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen & Richard N. Aslin. 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49(2), 85–117.

Pena, Marcela, Luca L. Bonatti, Marina Nespor & Jacques Mehler. 2002. Signal-driven computations in speech processing. *Science* 298, 604–607.

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

Petersson, Karl Magnus & Peter Hagoort. 2012. The neurobiology of syntax: beyond string sets. *Philosophical Transactions of the Royal Society B–Biological Sciences* 367, 1971–1983.

Petersson, Karl-Magnus, Vasiliki Folia & Peter Hagoort. 2012. What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language* 120(2), 83–95.

Poletiek, Fenna H. 2011. What in the world makes recursion so easy to learn? A statistical account of the staged input effect on learning a center-embedded structure in artificial grammar learning (AGL). *Biolinguistics* 5(1–2), 36–42.

Poletiek, Fenna H. & Jun Lai. 2012. How semantic biases in simple adjacencies affect learning a complex structure with non-adjacencies in AGL: A statistical account. *Philosophical Transactions of the Royal Society B–Biological Sciences* 367(1598), 2046–2054.

Radford, Andrew. 1988. *Transformational Grammar*. Cambridge: Cambridge University Press.

Rey, Arnaud, Pierre Perruchet & Joël Fagot. 2102. Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition* 123, 180–184.

Salomaa, Arto. 1969. On the index of a context-free grammar and language. *Information and Control* 14, 474–477.

Saito, Mamoru & Hajime Hoji. 1983. Weak crossover and Move-α in Japanese. *Natural Language and Linguistic Theory* 1, 245–259.

ten Cate, Carel & Kazuo Okanoya. 2012. Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B-Biological Sciences* 367(1598), 1984–1994.

Udden, Julia, Martin Ingvar, Peter Hagoort & Karl M. Petersson. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*

36(6), 1078–1101.

van Heijningen, Caroline A. A., Jos de Visser, Willem Zuidema & Carol ten Cate. 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America* 106(48), 20538–20543.

Yip, Moira J. 2006. The search for phonology in other species. *Trends in Cognitive Sciences* 10(10), 442–446.

Shiro Ojima
University of Tokyo
Department of Cognitive & Behavioral Sciences
Graduate School of Arts and Sciences
3-8-1 Komaba, Meguro-ku
Tokyo, 1538902
Japan
shiro.ojima@gmail.com

Kazuo Okanoya
University of Tokyo
Department of Cognitive & Behavioral Sciences
Graduate School of Arts and Sciences
3-8-1 Komaba, Meguro-ku
Tokyo, 1538902
Japan
kazuookanoya@gmail.com

Japan Science and Technology Agency
Okanoya Emotional Information Project, ERATO
Nibancho, Chiyoda-ku
Tokyo, 100-0004
Japan