Syntactic Capacities of Songbirds

Amir Kadivar

Abstract

Human language is, in many regards, unparalleled in the animal world. Although certain referential and computational aspects of language are reflected in non-human primates and songbirds, an evolutionary account for the emergence of the faculty of language remains elusive. In recent years, the capacity of songbirds to produce and recognize syntactic structures has gained much attention. All these experiments have focused on representative examples of artificial grammars inspired by Formal Language Theory and Automata Theory. Specifically, variations on two classes of artificial grammars, namely $(AB)^n$ *(a finite state grammar) and* A^nB^n *(a context-free grammar), have been used extensively to investigate the presence of recursive capacities in songbirds learn context free grammars. However, the highly idealized notion of recursion has proven difficult to test properly on human and non-human subjects. Furthermore, the posited central role of recursion in the development of human language has been disputed by some linguists. In light of these limitations and inspired by experimental evidence from human linguistics, we propose alternative artificial grammars that forgo the question of recursion and focus on a weaker, more tractable, capacity that is inarguably required for language processing, namely that of processing long-distance dependencies. We propose two experiments that investigate the ability of Bengalese Finches to learn and recognize the simplest possible long-distance dependency as well as the more complex crossed long-distance dependencies.*

1 Background

The parallels between human language acquisition and song learning in birds have been recognized at least from the time of Darwin [1]. Similar to human infants, song learning in birds shows a requirement for external input (tutoring), a critical developmental period ending with maturity, motor-auditory rehearsal systems, and hemispheric lateralization [2–6]. Furthermore, birdsong exhibits complex syntactic structures unparalleled by other non-human vocalization systems [5, 7–11].

Birdsong consists of stereotyped syllables (or notes) which the bird learns, through tutoring, to put in highly variable sequences the structure of which is species-specific. Different songbirds vary in their rates of song acquisition. Canaries, for instance, proceed to learn a new song every year. Zebra finches and Bengalese finches, in contrast, have a limited vocal learning period [7]. Deafening and auditory disruption experiments have shown that all songbirds show a reliance on sustained auditory feedback, even after song crystallization, to maintain song structure [6, 8].

In comparative studies of linguistic components in different animals one must distinguish between *auditory learning* and *vocal learning* [2, 12]. The former refers to the learned capacity to recognize auditory signals which is reflected in differential behavioral responses. For instance, songbirds recognize the song of other individuals and associate them with a limited range of possible meanings (conspecific or not). Similarly, a dog may learn to recognize the sound "sit" and to associate it with the act of sitting. Vocal learning, on the other hand, refers to the stronger capacity to not only recognize but immitate vocalizations. In this sense, almost all vertebrates are auditory learners while vocal learning has only been observed in three groups of mammals (humans, bats, and cetaceans) and three groups of birds (parrots, hummingbirds, and songbirds) [4, 12] implying a convergent evolutionary pathway for the development of certain aspects of language processing.

In this section we frame our central question regarding the capacity of songbirds to recognize and produce syntactic structures in the context of the broader question of the evolutionary development of the faculty of language.

1.1 Compnents and Counterparts of Human Language

The human language faculty, in its broadest sense, requires a wide variety of neural components ranging from sensorimotor systems, working and short-term memory, as well as higher mental faculties [13]. From a structural point of view, human languages are composed of a collection of rules organized conceptually as follows: *phonology* is the set of rules governing the concatenation of "sounds" (phonemes), *morphology* is the set of rules governing the concatenation of smallest meaningful units (morphemes) into words (for instance, *procedurally* is composed of three morphemes [[[*procedure*]-*al*]-*ly*]), and *syntax* is the set of rules governing the organization of words into phrases and sentences, most importantly word order. Probably the most formidable property of human language is its infinite capacity for creative production; the power to "make infinite use of finite means" [14]. The classic demonstration of this is the capacity of human language for arbitrarily long chains of nested subordinate clauses, a property known as *recursion* [2, 3, 13, 16–18].

Despite the conceptual convenience of this compartmentalization, these components are clearly not independent of each other nor are they meaningful without semantics. A *lexicon*, containing mappings of sounds to meanings, with a capacity to generalize individual sounds to syntactic categories (e.g. noun and verb) interacts with all these components. Due to all these interactions, it is clear that individual components of human language cannot be directly tested in non-human subjects. Instead, we need to identify fundamental computational processes that appear to be essential in certain aspects of language processing and seek neural counterparts in auditory sequence processing in other animals [2, 14, 15].

1.2 Syntactic Structure of Birdsong

Birdsong is typically highly variable in structure [5, 7, 8] and its complexity parallels the phonological complexity of human languages [2, 4, 16]. However, birdsong appears to serve only a narrow range of meaning, namely in the context of mate attraction, and lacks any semantic referential property. Despite elaborate variations in syllable order, birdsong appears to lack a higher level of structure beyond syllables (as opposed to syntactic structures in human language which exist beyond the level of phonology). As such, birdsong is considerd to have only a "phonological syntax" [2–4].

Nevertheless, this complex phonological structure requires parsing and sequencing neural machinery which may be analogous to the corresponding machinery involved in higher level syntactic processing in humans. Therefore, in the search for the neural substrate of syntactic processing, auditory learning in songbirds can serve as a model experimental system.

Pairwise transition probabilities between syllables are widely used to model the variability of birdsong [5, 7, 8, 10, 20] (Fig. 1), i.e as tools to capture the structure of birdsong as it relates to vocal learning (and not merely auditory learning). The underlying mathematical object in a Markov model is a random process defined by a collection of *states* and corresponding *transition probabilities*. The process starts at a designated *start state* and proceeds in each step by moving to a new state according to transition probabilities. At each new state one or more symbols are emitted. When the process arrives at a designated *stop state*, the sequence of symbols emitted is considered a valid vocalization in the language of interest. The defining feature of a Markov model is that transition probabilities between two states only depend on the two states and not on the history of the steps leading to the current state (i.e lack of *memory*).

1.3 Formal Language Theory and Automata Theory

Formal Language Theory (FLT) and Automata Theory (AT) are components of a mathematical theory of computation that underlies much of computer science. The general structure of the theory is as follows: FLT describes how languages can be represented using a *grammar* which is a collection of *production* (or *rewrite*) rules. For instance, the rewrite rule $S \rightarrow Sa|a$ defines the language $\{a^n\}$ consisting of all strings of arbitrary length containing the single symbol a.

On the other hand, AT describes automatons (abstract machines idealizing the notion of an *algorithm*) with a set of internal states and state transition rules. Automatons differ from Markov processes in that *1*) an automaton has no notion of transition probabilities and instead has a set of possible next states at each states (parallel nondeterminism), *2*) some automata are equipped with a memory component. Each automaton is capable of recognizing (or producing) a corresponding language.

Recognizing (or producing) exceedingly complex languages requires exceedingly complex automata. In this sense, languages and their corresponding automata are placed in a hierarchy of relative complexity [22] (Fig. 3). For instance, the automaton in Fig. 2 produces the same language $\{a^n\}$ as the grammar $S \rightarrow Sa|a$. Such machines with no memory component are called *finite-state automata* (FSA).

1.3.1 Regular and Context-Free Languages

Regular languages are those described by a grammar containing rules of the form $S \rightarrow Sa$. For instance the language $\{a^n\}$ discussed above is regular. The language $\{(ab)^n\} = \{ab, abab, ababab, \ldots\}$ is also regular as it is generated by the grammar $S \rightarrow Tb, T \rightarrow Sa$. It can be proved that a FSA recognizes regular languages and for any regular language there exists a FSA that recognizes it. FSAs and a regular languages are corresponding machine and language classes in the hierarchy.

Context-free languages (CFL) are those generated by context-free grammars (CFG) which lift the restriction imposed on regular grammars: the righthand side of a rewrite rule can be any combination of terminal or nonterminal symbols while the left handside still must only contain nonterminals. Consequently, regular languages are all necessarily CFL. An example of a CFL that is not regular is $\{a^n b^n\} = \{ab, aabb, aaabbb, \ldots\}$ defined by the grammar $S \rightarrow aSb$. Corresponding to CFLs, are a class of machines called *pushdown automata* (PDA) which are FSAs with an additional memory *stack* [1] such that transition rules between states of the FSA can depend on the value at the top of the stack and the automaton can add elements to the top of the stack over the course of its operation. It can be shown that PDAs recognize CFLs and all CFLs are recognized by some PDA and, therefore, CFLs and PDAs form another level of complexity.

1.4 Artificial Grammar Learning Experiments on Songbirds

As indicated above, vocalized birdsong appears to be no more complicated than what can be produced with a probabilistic FSA and thus birdsong is a regular language, insofar as vocal learning is concerned. A separate question, however, is whether the parsing machinery of songbirds can recognize more complex languages (i.e auditory learning). One paradigm to test such hypotheses is artificial grammar learning (AGL) where syntactically valid examples of a pre-designated language are presented to subjects in a familiarization period and their ability to distinguish between syntactically valid and invalid utterances is evaluated using a variety of behavioral measures [16, 23, 24]. In the past decade, AGL experiments have been fruitfully performed on humans [25], primates [26], and songbirds [27, 28]. With an eye towards FLT and AT, regular and context-free languages have been used as examples in such experiments. One study shows that non-human primates fail to learn to recognize context-free languages [23], and others show that songbirds have such a capacity [27, 28]. In this section we outline the known methodological flaws in birdsong AGL experiments with CFL examples. In the next section, with the broader evolutionary question, we outline the conceptual flaws in the same experiments.

In the first CFL experiment in birdsong [27], European starlings were familiarized with examples of the regular language $\{(ab)^n\}$ and the CFL $\{a^n b^n\}$ where a and b are two of the species' syllables. The increased call rate of the subjects in the test stage to grammatical utterances is interpreted as indicative of the capacity of songbirds to recognize context-free languages. There are two critical flaws with this experiment design.

---

[1] A *stack* is a list together with two read and write operations: *push* adds an item to the top of the list, and *pop* removes and reads an item from the top of the list.

First, a more parsimonious explanation for the recognition capacity is a count-and-compare strategy where the birds count the number of occurances of the a syllable and compare it with that of the b syllable. [2] Furthermore, although $\{a^n b^n\}$ is a CFL and human languages are mostly CFLs, $\{a^n b^n\}$ is a form that appears in no human language and even human subjects are likely to be employing the same count-and-compare strategy [2, 25, 29]. Second, for this example CFL to have any bearing on the capacity to parse hierarchical syntactic structures a one to one and nested correspondance between the a syllables and b syllables must be established. This is impossible to infer from the experiment where all a syllables are identical. A solution is to extend the a symbol to contain a phonetic category (a collection of syllables instead of a single syllable) which is paired with syllables of a second b category. In the modified language example sequences have the form $a_1 a_2 \dots a_n b_n \dots b_2 b_1$.

In the second CFL experiment in birdsong [28], Bengalese finches are familiarized with examples of the modified CFL discussed above. First, all available syllables are arbitrarily classified in three categories A, C, F. Second, birds are familiarized with sequences from the language $\{A_i A_j C_k F_j F_i\}$ which captures the nested pairing of syllables in addition to the condition that $k \neq i, j$ The results, again, indicated that birds responded differentially to syntactically valid sequences. However, again, an experiment design flaw brings the results under question [29] (Fig. 4). with a more parsimonious explanation that birds learn to generalize the middle syllable and use the usual syllable transition probabilities to recognize syntactically valid sequences.

1.5 Applicability of FLT and AT to Neural Systems

There are various deep differences between an automaton model of computation and the neural network structure of the brain that limits the applicabilty of FLT and AT to neural circuits [15, 23]. Here, we note three main idealizing assumptions of FLT and AT that are violated in physiological systems.

First, a crucial feature of automata (FSA, PDA, or more complex machines) is their nondeterminism in the sense that each state can have multiple outgoing transitions to other states. For finite state automata it can be shown that deterministic automata (DFA) and nondeterministic ones (NFA) are essentially equally complex in the sense that they recognize the same class of languages, namely regular languages. In the case of pushdown automata however, nondeterminism is an essential feature (i.e deterministic PDAs cannot recognize the full CFL class) [22]. Alhtough nondeterminism is well reflected in the Markov models of vocalized song syntax and the physiological neural network models underlying it, the abstract machine nondeteremInism requires that at each state multiple copies of the machine are launched following distinct transition trajectories [22]. This contradicts physiological requirements hypothesized for branch/chain network models of song production and sequencing [9].

Second, an important feature of a language that can render it more or less complex is the unboundedness of its strings. For instance, the CFL $\{a^n b^n\}$ is not regular however $\{a^n b^n; n \leq N\}$ is regular for any N (e.g.

---

[2] That songbirds along with many other animals, including most primates, have the capacity to count and compare is well established [2, 23, 28, 29].

$\{ab, aabb, aaabbb\}$ is regular and can be recognized by a FSA [23].

Finally, The abstraction of the memory device in AT has certain pecularities that are presumably not reflected in brain structure. First, the requirement that the memory of a PDA be a single stack (and not other forms of a list) is crucial. For instance, if the single stack is replaced by a queue or even by two stacks, the PDA is suddently transformed into a full-blown Turing machine; the most powerful universal machine. This has dubious consequences if taken seriously physiologically. For instance, assuming there is a neural substrate opereating like a stack there is a strong discrepancy between the miniscule biological cost of adding a second stack and the theoretical explosion of expressive power.

2 Experiments

In light of the above limitations, as well as the dispute among linguists regarding the centrality of recursion in the language faculty [13, 17, 30–32], we propose to forgo the question of placing songbird capacities in the language hierarchy (Fig. 3). Instead, we focus on a specific computational requirement, namely the parsing memory, which is inarguably involved in human language processing. This computational requirement is equivalent, from the standpoint of FLT and AT, to *supra-regularity*. A supra-regular language is one that ranks higher in the hierarchy than regular langugaes which are recognized by memoryless automata (i.e FSAs). However, instead of considering the usual $\{a^n b^n\}$ CFG, with its interpretation as a language with nested long-distance dependencies, we consider simpler supra-regular grammars that involve the simplest form of long-distance dependencies.

2.1 Methods

All the following AGL experiments follow the same methodology described in earlier experiments [26–28]. Syllables (henceforth refered to as $A, B, C, \ldots$) from a corpus of Bengalese finch song will be selected. Thesse syllables are used to build the artificial grammars and the corresponding familiarization and test strings described below. Subject birds will be exposed to grammar-conforming familiarization strings of each language over 60 minutes. To test grammar acquisition, the change in the call response number over a 5 minute period before and after the exposure to familiarization sequences will be analyzed. A significant behavioral difference in response to novel grammar-conforming strings as compared to grammar-violating strings after the familiarization period would provide an indication of artificial grammar acquisition. In accordance with earlier studies, we expect the behavioral response to ungrammatical sequences to involve birds ceasing their movements and reducing their basal calling rates [28].

2.2 Experiment I: Long-Distance Dependencies

In this experiment we consider the simplest context-free grammar containing long-distance dependencies: $\{xS_1x, xS_2x, xS_3x, \ldots\}$ where x is either of A or B and S_i are arbitrary filler strings not containing A or

B (Fig. 5). This grammar has multiple advantages over those used in earlier experiments:

- The only grammatical category that birds need to learn is the special role of A and B syllables (as opposed to the more complicated categories of earlier experiments [28]).
- The S_i sequences can be arbitrarily chosen and reused both in familiarization and test strings. This allows for a natural control for any possible strategy that would involve the acoustic properties of these filler sequences.
- The simplicity of the grammar allows for a large number of sequences to be used for familiarization and testing blocking the possibility of memorization [2].
- No local transition rule can capture the long-distance dependency since any starting subsequence xS_i or ending subsequence S_ix can be grammatical or ungrammatical.

An example set of familiarization and test sequences using only five syllables A, B, C, D, E is as follows. The set of filler sequences

$$F = \{CDE, CED, DCE, DEC, ECD, EDC\}$$

is randomly split into two groups F_1 and F_2 for each subject. Sequences of the form AS_iA and BS_iB where $S_i \in F_1$ are chosen for familiarization. Four classes of test sequences are produced as follows: Sequences of the form AS_iB and BS_iA where $S_i \in F_1$ are chosen as ungrammatical test sequences with familiar filler sequences. Sequences of the form AS_iB and BS_iA where $S_i \in F_2$ are chosen as ungrammatical sequences with novel filler sequences. Similarly, sequences of the form AS_iA and BS_iB for either $S_i \in F_1$ or $S_i \in F_2$ are chosen as grammatical test strings with familiar or novel filler sequences. A similar procedure can be applied to larger filler sets F using a larger number of syllables (there are up to 16 distinct syllables in Bengalese Finch songs [28]) to provide larger familiarization and test sequences with similar properties.

2.3 Experiment II: Crossed Dependencies and Context-Sensitive Languages

In this experiment we try to show how the failures in the idealized assumptions of FLT and AT can lead to surprising results. We present the subjects to syntactic examples of *context-sensitive* languages (which, in general, are arguably only parsable by humans). Context-sensitive languages (CSL) are those generated by context-sensitive grammars (CSG) in which the restriction on the lefthand side of rewrite rules is lifted. In a CSG the lefthand side too can be any combination of terminal or nonterminal symbols, e.g. $aS \to aSc$ and $bS \to bSb$ which cannot be expressed in a CFG.

First, to shed more light on the applicability of FLT and AT on animal grammar learning, we propose to test a simple context-sensitive grammar which is theoretically much more complex than a context-free grammar but can be recognized, in principle, with a count and compare strategy as discussed above. The artificial grammar of interest is:

$$\{x^ny^nz^n\} = \{xyz, xxyyzz, xxxyyyzzz, \ldots\}$$

The strings can be fixed to a specific length to simplify the learning task (e.g. only $n = 2$). Then various orderings of the four syllables A, B, C, D can be used for familiarization and test. For instance, the familiarization set would be:

$$\{AABBCC, AACCDD, AABBDD, BBCCDD\}$$

with grammatical test strings:

$$\{DDCCBB, DDCCAA, DDBBAA, CCBBAA\}$$

and ungrammatical test strings:

$$\{AABCBC, AACDCD, AABDBD, BBCDCD, CCABAB, DDACAC\}$$

$$\{ABCCDD, ACBBDD, ADBBCC, BCAADD, CDAABB, CDBBAA\}$$

Second, we consider the *copy* language: $\{\omega\omega\}$ where ω is any arbitrary sequence. This language is another context-sensitive grammar reflecting a more complex form of long-distance dependency known as *crossed long-distance* (or *cross-serial*) dependencies found in human languages such as Dutch [15] (Fig. 6, 7). This language, too, could in principle be recognized for limited sequence sizes, by a comparison with a working memory copy of the first occurence of ω. Furthermore, it has been shown that a strong distinction exists between two and three crossed long-distance dependencies in terms of their ability to be parsed by humans [15].

Similar to the first CSG, we can fix the length of ω, say to 3, and consider the set W of all 27 possible 3 symbol sequences. Similar to experiment I, W will be split randomly for each subject to W_1 and W_2 with familiarization sequences $\omega\omega$ chosen from $\omega \in W_1$ and W_2 is reserved for novel grammatical test strings. Ungrammatical test strings $\omega_1\omega_2$ can be arbitrarily chosen for $\omega_1, \omega_2 \in W$. The occurence of all possible local transitions in W (and thus in W_1 and W_2) provides a natural control for any learning strategy that requires such transition probabilities.

References

1. Darwin, C. *The descent of man, and selection in relation to sex* (Murray, 1888).
2. Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* **17,** 89–98 (2013).
3. Miyagawa, S., Berwick, R. C. & Okanoya, K. The emergence of hierarchical structure in human language. *Frontiers in Psychology* **4,** 71 (2013).
4. Bolhuis, J. J., Okanoya, K. & Scharff, C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* **11,** 747–759 (2010).

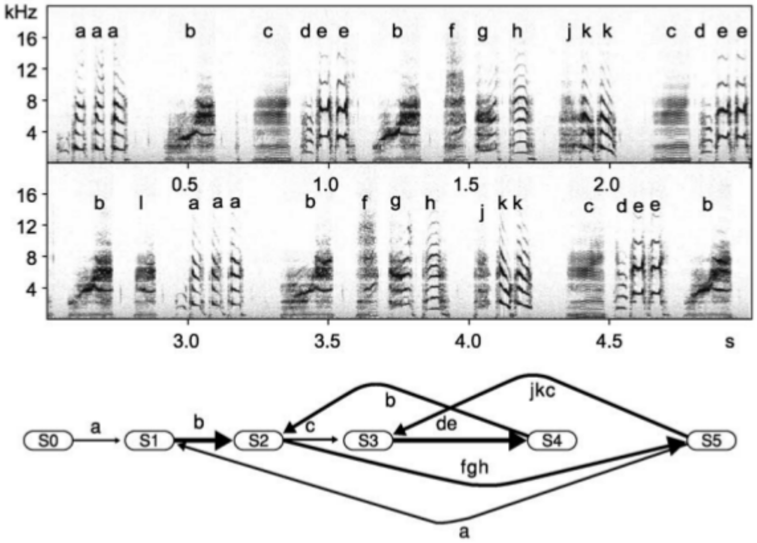

Figure 1: Markov model for Bengalese finch song (probabilities not shown; taken from [7]).

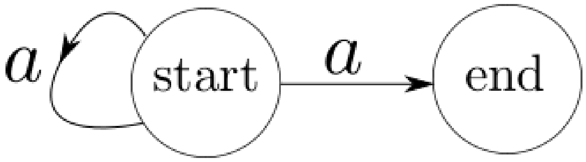

Figure 2: Finite State Automaton that recognizes the language $\{a^n\}$.

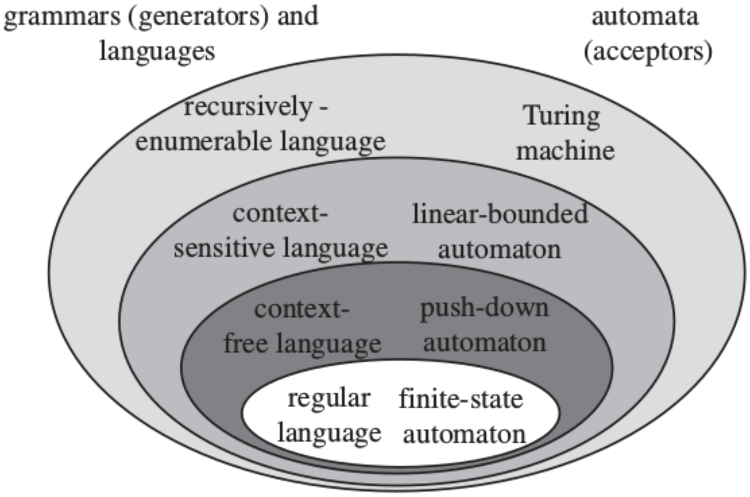

Figure 3: The Chomsky hierarchy of language classes and their corresponding automata (taken from [23]).

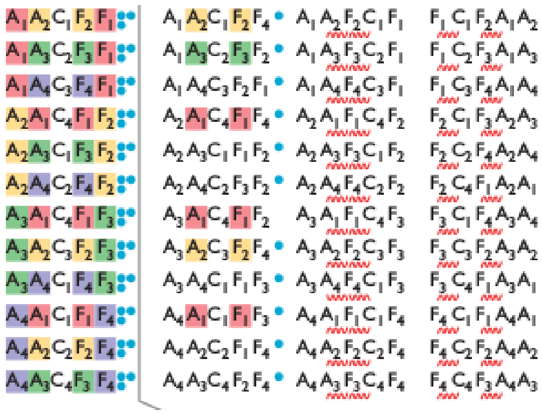

Figure 4: An example showing overrepresentation of syntactic tests (left) when compared to syntax-violating tests (right) in familiarization strings. Blue dots represent the number of familiarization strings that match the C-generalized test string and red underlines show transition probability violations when compared to familiarization strings (taken from [29]).

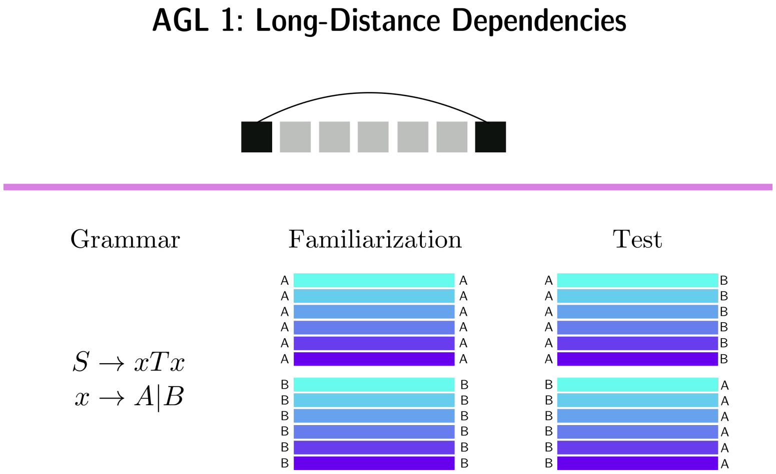

Figure 5: Schematic description of the first AGL grammar containing the simplest form of long-distance dependency. The filler sequences S_i are color coded with different shades.

Figure 6: Example of crossed long-distance dependencies requiring a context-sensitive grammar.

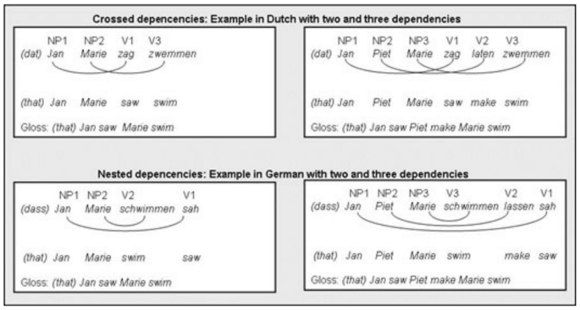

Figure 7: Example of crossed long-distance dependencies in Dutch as compared to nested (context-free) dependencies in German (taken from [15]).

5. Okanoya, K. & Yamaguchi, A. Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. *Journal of Neurobiology* **33,** 343–356 (1997).
6. Sakata, J. T. & Brainard, M. S. Real-time contributions of auditory feedback to avian vocal motor control. *The Journal of Neuroscience* **26,** 9619–9628 (2006).
7. Okanoya, K. The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences* **1016,** 724–735 (2004).
8. Okanoya, K. Song syntax in Bengalese finches: proximate and ultimate analyses. *Advances in the Study of Behavior* **34,** 297–346 (2004).
9. Jin, D. Z. Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Physical Review E* **80,** 051902 (2009).
10. Jin, D. Z. & Kozhevnikov, A. A. A compact statistical model of the song syntax in Bengalese finch. *PLoS Computational Biology* **7,** e1001108 (2011).
11. Katahira, K., Suzuki, K., Okanoya, K. & Okada, M. Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PloS ONE* **6,** e24516 (2011).
12. Jarvis, E. D. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* **1016,** 749–777 (2004).
13. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298,** 1569–1579 (2002).
14. Chomsky, N. *Cartesian linguistics: A chapter in the history of rationalist thought* (Cambridge University Press, 2009).
15. De Vries, M., Christiansen, M. H. & Petersson, K. M. Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* **5,** 010–035 (2011).
16. Ten Cate, C. & Okanoya, K. Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **367,** 1984–1994 (2012).
17. Pinker, S. & Jackendoff, R. The faculty of language: what's special about it? *Cognition* **95,** 201–236 (2005).

18. Petkov, C. I. & Wilson, B. On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367,** 2077–2088 (2012).

19. Fujimoto, H., Hasegawa, T. & Watanabe, D. Neural coding of syntactic structure in learned vocalizations in the songbird. *The Journal of Neuroscience* **31,** 10023–10033 (2011).

20. Hanuschkin, A., Diesmann, M. & Morrison, A. A reafferent and feed-forward model of song syntax generation in the Bengalese finch. *Journal of Computational Neuroscience* **31,** 509–532 (2011).

21. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77,** 257–286 (1989).

22. Sipser, M. Introduction to the theory of computation. *ACM SIGACT News* **27,** 27–29 (1996).

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

24. Ten Cate, C. On the phonetic and syntactic processing abilities of birds: from songs to speech and artificial grammars. *Current Opinion in Neurobiology* **28,** 157–164 (2014).

25. Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I. & Anwander, A. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 2458–2463 (2006).

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

27. Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. Recursive syntactic pattern learning by songbirds. *Nature* **440,** 1204–1207 (2006).

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

29. Beckers, G. J., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23,** 139–145 (2012).

30. Fitch, W. T., Hauser, M. D. & Chomsky, N. The evolution of the language faculty: clarifications and implications. *Cognition* **97,** 179–210 (2005).

31. Jackendoff, R. Possible stages in the evolution of the language capacity. *Trends in Cognitive Cciences* **3,** 272–279 (1999).

32. Jackendoff, R. & Wittenberg, E. Linear grammar as a possible stepping-stone in the evolution of language. *Psychonomic Bulletin & Review,* 1–6 (2016).

33. Miyagawa, S., Ojima, S., Berwick, R. C. & Okanoya, K. The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in Psychology* **5,** 564 (2014).