Review

Cell PRESS

Songs to syntax: the linguistics of birdsong

Robert C. Berwick[1], Kazuo Okanoya[2,3], Gabriel J.L. Beckers[4] and Johan J. Bolhuis[5]

[1] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Cognitive and Behavioral Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
[3] RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-City, Saitama 351-0198, Japan
[4] Department of Behavioural Neurobiology, Max Planck Institute for Ornithology, D-82319 Seewiesen, Germany
[5] Behavioural Biology and Helmholtz Institute, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Unlike our primate cousins, many species of bird share with humans a capacity for vocal learning, a crucial factor in speech acquisition. There are striking behavioural, neural and genetic similarities between auditory-vocal learning in birds and human infants. Recently, the linguistic parallels between birdsong and spoken language have begun to be investigated. Although both birdsong and human language are hierarchically organized according to particular syntactic constraints, birdsong structure is best characterized as 'phonological syntax', resembling aspects of human sound structure. Crucially, birdsong lacks semantics and words. Formal language and linguistic analysis remains essential for the proper characterization of birdsong as a model system for human speech and language, and for the study of the brain and cognition evolution.

Human language and birdsong: the biological perspective

Darwin [1] noted strong similarities between the ways that human infants learn to speak and birds learn to sing. This 'perspective from organismal biology' [2] initially led to a focus on apes as model systems for human speech and language (see Glossary), with limited success, however [3,4]. Since the end of the 20th century, biologists and linguists have shown a renewed interest in songbirds, revealing fascinating similarities between birdsong and human speech at the behavioural, neural, genomic and cognitive levels [5–9]. Yip has reviewed the relationship between human phonology and birdsong [7]. Here, we address another potential parallel between birdsong and human language: syntax.

Comparing syntactic ability across birds and humans is important, because at least since the beginning of the modern era in cognitive science and linguistics, a combinatorial syntax has been viewed to lie at the heart of the distinctive creative and open-ended nature of human language [10]. Here, we discuss current understanding of the relationship between birdsong and human syntax in light of recent experimental and linguistic advances, focusing on the formal parallels and their implications for underlying cognitive and computational abilities. Finally, we sketch the prospects for future experimental work, as part of the

Glossary

Bigram: a subsequence of two elements (notes, words or phrases) in a string.
Context-free language (CFL): the sets of strings that can be recognized or generated by a pushdown-stack automaton or context-free grammar. A CFL might have grammatical dependencies nested inside to any depth, but dependencies cannot overlap.
Finite-state automaton (FSA, FA): a computational model of a machine with finite memory, consisting of a finite set of states, a start state, an input alphabet, and a transition function that maps input symbols and current states to some set of next states.
Finite-state grammar (FSG): a grammar that formally replicates the structure of a FSA, also generating the regular languages.
***K*-reversible finite-state automaton**: an FSA that is deterministic when one 'reverses' all the transitions so that the automaton runs backwards. One can 'look behind' *k* previous words to resolve any possible ambiguity about which next state to move to.
Language: any possible set of strings over some (usually finite) alphabet of words.
Locally testable language: a strict subset of the regular languages formed by the union, intersection, or complement of strictly locally testable languages.
(First-order) Markov model or process: a random process where the next state of a system depends only on the current state and not its previous states. Applied to word or acoustic sequences, the next word or acoustic unit in the sequence depends only on the current word or acoustic unit, rather than previous words or units.
Mildly context-sensitive language (MCSL): a language family that lies 'just beyond' the CFLs in terms of power, and thought to encompass all the known human languages. A MCSL is distinguished from a CFL in that it contains clauses that can be nested inside clauses arbitrarily deeply, with a limited number of overlapping grammatical dependencies.
Morphology: the possible 'word shapes' in a language; that is, the syntax of words and word parts.
Phoneme: the smallest possible meaningful unit of sound.
Phonetics: the study of the actual speech sounds of all languages, including their physical properties, the way they are perceived and the way in which vocal organs produce sounds.
Phonology: the study of the abstract sound patterns of a particular language, usually according to some system of rules.
Push-down stack automaton (PDA): a FSA augmented with a potentially unbounded memory store, a push-down stack, that can be accessed in terms of a last-in, first-out basis, similar to a stack of dinner plates, with the last element placed on the stack being the top of the stack, and first accessible memory element. PDAs recognize the class of CFLs.
Recursion: a property of a (set of) grammar rules such that a phrase *A* can eventually be rewritten as itself with non-empty strings of words or phrase names on either side in the form $\alpha A \beta$ and where *A* derives one or more words in the language.
Regular language: a language recognized or generated by a FSA or a FSG.
Semantics: the analysis of the meaning of a language, at the word, phrase, sentence level, or beyond.
Strictly locally testable language (or stringset): a strict subset of the regular languages defined in terms of a finite list of strings of length less than or equal to some upper length *k* (the 'window length').
Sub-regular language: any subset of the regular languages, in particular generally a strict subset with some property of interest, such as local testability.
Syllable: in linguistics, a vowel plus one or more preceding or following consonants.
Syntax: the rules for arranging items (sounds, words, word parts or phrases) into their possible permissible combinations in a language.

Corresponding author: Bolhuis, J.J. (j.j.bolhuis@uu.nl).

1364-6613/$ – see front matter doi:10.1016/j.tics.2011.01.002 Trends in Cognitive Sciences xx (2011) 1–9

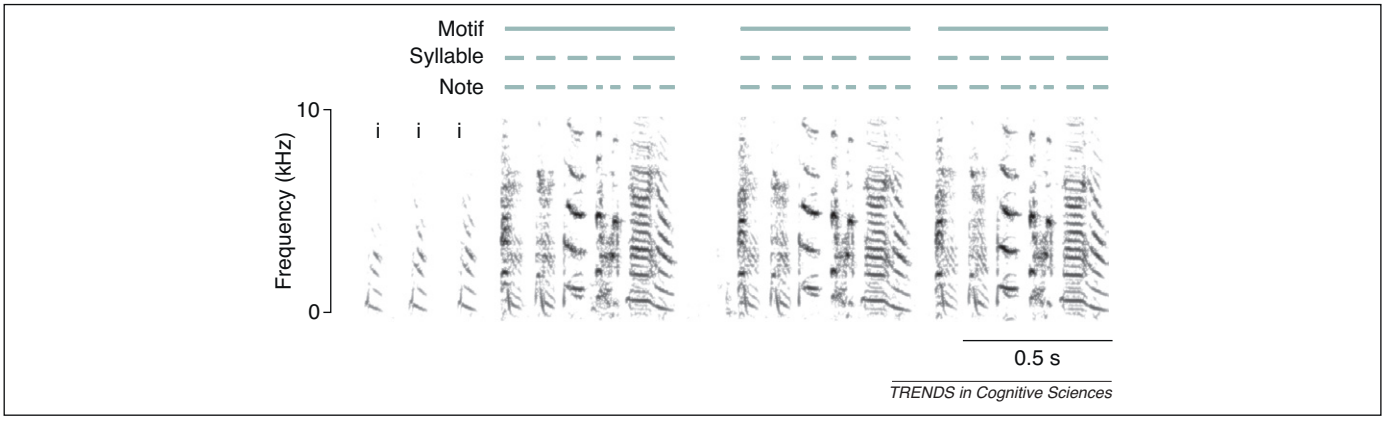

Figure 1. Sound spectrogram of a typical zebra finch song depicting a hierarchical structure. Songs often start with 'introductory notes' (denoted by 'i') that are followed by one or more 'motifs', which are repeated sequences of syllables. A 'syllable' is an uninterrupted sound, which consists of one or more coherent time-frequency traces, which are called 'notes'. A continuous rendition of several motifs is referred to as a 'song bout'.

ongoing debate as to what is species specific about human language [3,11]. We show that, although it has a simple syntactic structure, birdsong cannot be directly compared with the syntactic complexity of human language, principally because it has neither semantics nor a lexicon.

Comparing human language and birdsong

Human speech and birdsong both consist of complex, patterned vocalizations (Figure 1). Such sequential structures can be analysed and compared via formal syntactic methods. Aristotle described language as sound paired with meaning [12]. Although partly accurate, a proper interspecies comparison calls for a more articulated 'system diagram' of the key components of human language, and their non-human counterparts. We depict these as a tripartite division (Figure 2): (i) an 'external interface', a sensorimotor-driven, input–output system providing proper articulatory output and perceptual analysis; (ii) a rule system generating correctly structured sentence forms, incorporating words; and (iii) an 'internal interface' to a conceptual–intentional system of meaning and reasoning; that is, 'semantics'. Component (i) corresponds to systems for producing, perceiving and learning acoustic sequences, and might itself involve abstract representations that are not strictly sensorimotor, such as stress placement. In current linguistic frameworks, (i) aligns with acoustic phonetics and phonology, for both production and perception. Component (ii) feeds into both the sensorimotor interface (i), as well as a conceptual–intentional system (iii), and is usually described via some model of recursive syntax.

Although linguists debate the details of these components, there seems to be more general agreement as to the nature of (i), less agreement as to the nature of (ii) and widespread controversy as to (iii). For instance, whereas the connection between a fully recursive syntax and a conceptual–intentional system is sometimes considered to lie at the heart of the species-specific properties of human language, there is considerable debate over the details, which plays out as the distinct variants of current linguistic theories [13–16]. Some of these accounts reduce or even eliminate the role of (ii), assuming a more direct relation between (i) and (iii) (e.g. [17,18]). The system diagram in Figure 2 therefore cannot represent any detailed neuroanatomical or abstract 'wiring diagram', but

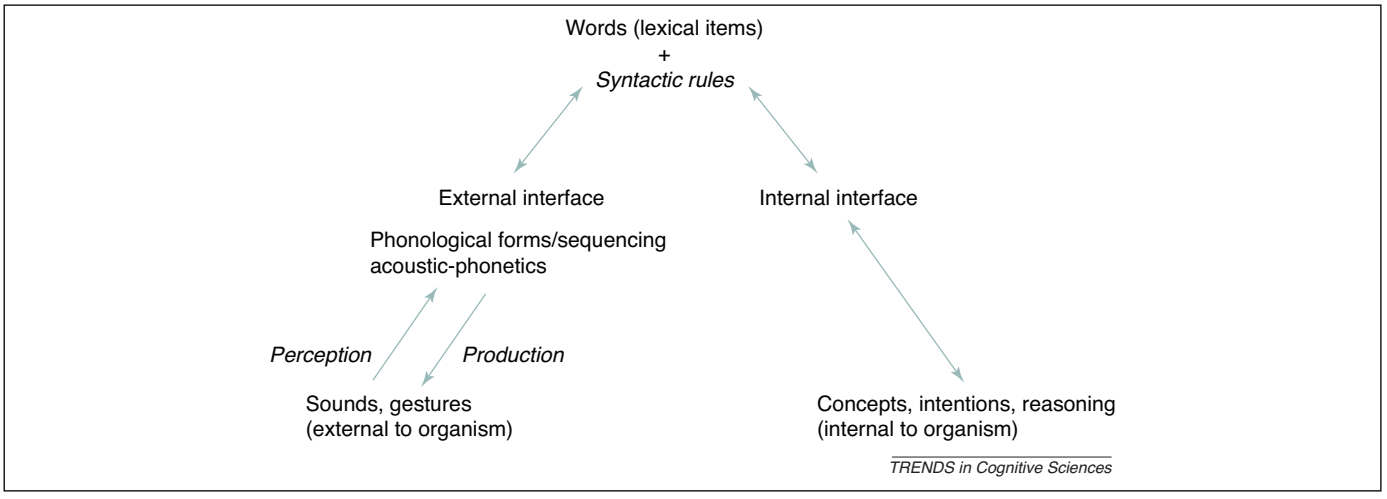

Figure 2. A tripartite diagram of abstract components encompassing both human language and birdsong. On the left-hand side, an external interface **(i)**, comprised of sensorimotor systems, links the perception and production of acoustic signals to an internal system of syntactic rules, **(ii)**. On the right-hand side, an internal interface links syntactic forms to some system of concepts and intentions, **(iii)**. With respect to this decomposition, birdsong seems distinct from human language in the sense of lacking both words and a fully developed conceptual–intentional system.

rather a way to factor apart the distinct knowledge types in the sense of Marr [19]. Notably, our tripartite arrangement does not preclude the possibility that only humans have syntactic rules, or that such rules always fix information content in a language-like manner. For example, in songbirds, sequential syntactic rules might exist only to construct variable song element sequences rather than variable meanings *per se* [9].

Birdsong and human syntax: similarities and differences

Both birdsong and human language are hierarchically organized according to syntactic constraints. We compare them by first considering the complexity of their sound structure, and then turning in the next section, to aspects beyond this dimension. Overall, we find that birdsong sound structure, at least for the Bengalese finch, seems characterizable by an easily learnable, highly restricted subclass of the regular languages (languages that can be recognized or generated by finite-state machines; see Box 3). Whereas human language sound structure also appears to be describable via finite-state machines, comparable results are lacking in the case of human language, although certain parts of human language sound structure, such as stress patterns, have also recently been shown to be easily learnable [20].

In birdsong, individual notes can be combined as particular sequences into syllables, syllables into 'motifs', and motifs into complete song 'bouts' (Figure 1). Birdsong thus consists of chains of discrete acoustic elements arranged in a particular temporal order [21–23]. Songs might consist of fixed sequences with only sporadic variation (e.g. zebra finches), or more variable sequences (e.g. nightingales, starlings, or Bengalese finches), where a song element might be followed by several alternatives, with overall song structure describable by probabilistic rules between a finite number of states [23,24] (Figure I, Box 1). For example, a song of a nightingale is built out of a fixed 4-second note sequence. An individual nightingale has 100–200 song types, clustered into 2–12 'packages'. Package singing order remains probabilistic [25]. A starling song bout might last up to 1 minute, composed of many distinct motifs containing song elements in a fixed order lasting 0.5–1.5 seconds. Gentner and Hulse [26] found that a first-order Markov model (i.e. bigrams) suffices to describe most motif sequence information in starling songs (Box 2). Thus, for the most part, the next motif is predictable by the immediately preceding motif. Starlings also use this information to recognize specific song bouts. Similarly, in American thrush species, relatively low-order Markov chains suffice for modelling song sequence variability [27].

Can songbird 'phonological syntax' [28] ever be more complex than this? Bengalese finch song typically contains approximately eight song note types organized into 2–5 note 'chunks' that also follow local transition probabilities [29] (Figure I, Box 1). Unlike single-note Markov processes, chunks such as the three-note sequence *cde* can be reused in other places in a song [24,30]. However, chunks are not reused inside other chunks, so the hierarchical depth is strictly limited.

If Bengalese finch song could be characterized solely in terms of bigrams, it would belong to the class of so-called 'strictly locally 2-testable languages', a highly restricted subset of the class of the regular languages, That is, a bird could verify, either for purposes of production or for recognition, whether a song is properly formed by simply 'sliding' a set of two-note sequences or 'window constraints' across the entire note sequence, checking to see that all the two-note sequences found 'pass' (Box 3). For example, if the valid note sequences were *ab, abab, ababab,* and so on, then every *a* must be followed by a *b*, except at the song start; and every *b* must be followed by an *a,* except at the song end. Thus, aside from the beginning and end of a song, a bird could check whether a song is well formed by using two bigram templates: [*a*-*b*] and [*b*-*a*]. This turns out to be the simplest kind of pattern recognizable by a finite-state automaton (FSA), because the internal states of the automaton need not be used for any detailed computation aside from bigram note template matching (Box 3).

The Bengalese finch song automaton in Figure I (Box 1), which encompasses the full song sequence repertoire extracted from a single, actual bird [31], indicates that birdsong structure can be more complicated than a simple

Box 1. Birdsong, human language syntax and the Chomsky hierarchy

All sets of strings, or languages, can be rank ordered via strict set-inclusion according to their computational power. The resulting 'rings' are called the 'Chomsky hierarchy' [61] (Figure I; ring numbers are used below). For birdsong and human syntax comparisons, the most important point is the small overlap between the possible languages generated by human syntax (the irregular-shaded grey set), as opposed to birdsong syntax (the stippled grey set).

1. The finite languages, all sets of strings of finite length.
2. The FSA generating the regular languages. An FSA is represented as a directed graph of states with labelled edges, a finite-state transition network. The corresponding grammar of an FSA has rules of the form $X \rightarrow aY$ or $X \rightarrow a$, or right-linear, where X and Y range over possible automaton states (nonterminals), and a ranges over symbols corresponding to the labelled transitions between states. The FSA recognizing the $(ab)^1$ language only need to test for four specific adjacent string symbol pairs (bigrams; the pairs (left-edge, *a*); (*a*,*b*); (*b, a*); and (*b,* right-edge) [62].
3. The PDA, generating the CFLs. PDAs are finite-state machines augmented with a potentially unbounded auxiliary memory that can be accessed from the top working down. PDAs can be thought of as augmenting FSA with the ability to use subroutines, yielding the recursive transition networks. Grammars for these languages are consequently more general and can include rules such as $X \rightarrow Ya$, $X \rightarrow aYa$ or $X \rightarrow aXa$, or context-free rules.
4. The PDA whose stacks might themselves be augmented with embedded stacks, generating the MCSLs. Examples of such patterns in human languages are rare, but do exist [63,64]. These patterns are exemplified by stringsets such as $a^n b^m c^n d^m$, where the *a*s and *c*s must match up in number and order and, separately, the *b*s and the *c*s, so-called 'cross-serial' dependencies (see [65,66]). A broad range of linguistic theories accommodate this much complexity [13–16,59,66]. No known human languages require more power than this. The two irregular sets drawn cutting across the hierarchy depict the probable location of the human languages (shaded) and birdsong (stippled). Both clearly do not completely subsume any of the previously mentioned stringsets. Birdsong and human languages intersect at the very bottom owing to the possible overlap of finite lists of human words and the vocal repertoire of certain birds.

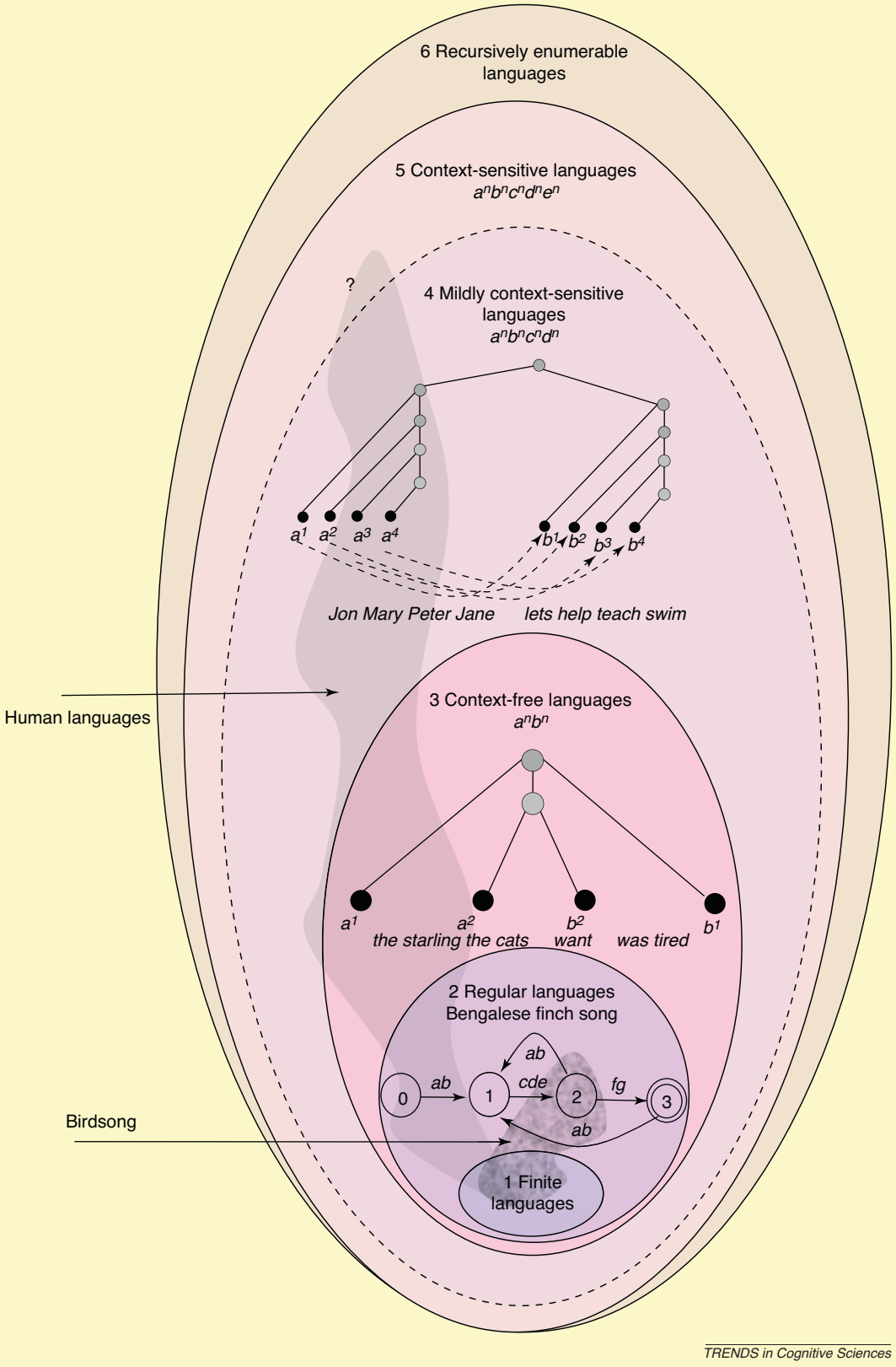

Figure I. The Chomsky hierarchy of languages along with the hypothesized locations of both human language and birdsong. The nested rings in the figure correspond to the increasingly larger sets, or languages, generated or recognized by more powerful automata or grammars. An example of the state transition diagram corresponding to a typical Bengalese finch song [31] is shown in the next ring after this, corresponding to some subset of the regular languages.

Box 2. Is recursion for the birds?

Recursive constructs occur in many familiar human language examples, such as *the starling the cats want was tired,* where one finds a full sentence phrase, *the starling was tired*, that contains within it a second, 'nested' or 'self-embedded' sentence, S, *the cats want*. In this case, the rule that constructs Sentences can apply to its own output, recursively generating a pattern of 'nested' or 'serial' dependencies.

We can write a simple CFG with three rules that illustrates this concept as follows: $S \rightarrow aB$; $B \rightarrow Sb$; $S \rightarrow \varepsilon$, where ε corresponds to the empty symbol. We can use this grammar to show that one can first apply the rule that expands *S* as *aB* and then can apply the second rule to expand *B* as *Sb,* thus obtaining, *aSb*; *S* now appears with non-null elements on both sides, so we say that *S* has been 'self-embedded'. If we now use the third rule to replace *S* with the empty symbol, we obtain the output *ab*. Alternatively, we could apply the first and second rules over again to obtain the string *aabb*, or, more generally, $a^n b^n$ for any integer *n*.

In our example, the *a*s and the *b*s in fact form nested dependencies because they are correspondingly paired in the same way that *the starling* must be paired with the singular form *was*, rather than the plural *were*; similarly, *the cats* must be paired with *want* rather than the singular form *wants.* So, for example, to indicate a nested dependency pattern properly, the form a^3b^3 should be more accurately written as $a^1a^2a^3b^3b^2b^1$, where the superscripts indicate which *a*s and *b*s must be paired up. Thus, any method to detect whether an animal can either recognize or produce a strictly context-free pattern requires that one demonstrates that the correct *a*s and *b*s are paired up; merely recognizing that the number of *a*s matches the number of *b*s does *not* suffice. This is one key difficulty with the Gentner *et al.* protocol and result [56], which probed only for the ability of starlings to recognize an equal number of *a*s and *b*s in terms of *warble* and *rattle* sound classes (i.e. $warlble^3rattle^3$ patterns) but did not test for whether these *warble-rattles* were properly paired off in a nested dependency fashion. As a result, considerable controversy remains as to whether any non-human species can truly recognize strictly context-free patterns [11,67].

Box 3. Descriptive complexity, birdsong and human syntax

The substructure of the regular languages, sub-regular language hierarchies, could be relevant to gain insight into the computational capacities of animals and humans in the domain of acoustic and artificial language learning [62,68,69]. Similar to the Chomsky hierarchy, the family of regular languages can itself be ordered in terms of strictly inclusive sets of increasing complexity [69]. The ordering uses the notion of descriptive complexity, corresponding informally to how much local context and internal state information must be used by a finite-state machine to recognize a particular string pattern correctly. For example, to recognize the regular pattern used in the starling experiment [56], $(ab)^1$, a finite-state machine needs only to check four adjacency relations or bigrams as they appear directly in a candidate string: the beginning of the string followed by an *a*; an *a* followed by a *b*;, a *b* followed by an *a* or else a *b* followed by the end of the string. We can say such a pattern is strictly locally 2-testable or SL_2 [69]. As we increase the length of these factors, we obtain a strictly increasing set hierarchy of regular languages, the strictly locally testable languages, denoted SL_k, where *k* is the 'window length' [56,62,68]. It might be of some value to understand the range of sub-regular patterns that birds can perceive or produce. To tentatively answer this question, we applied a program for computing local testability [38,44,70]. For example, the FSA in Figure I (Box 1) recognizes a language that is locally testable. This answer agrees with the independent findings of Okanoya [31] and Gentner [26,57].

Other sub-regular pattern families have been recently explored in connection with human language sound systems [20,71]. Some of these might ultimately prove relevant to birdsong because they deal with acoustic patterns. In particular, possible sound combinations might fall into the same classes as those of human languages. Finally, all these sub-regular families could be extended straightforwardly to include phrases explicitly, but still without the ability to 'count', as seems true of human language ([66,72–74] R. Berwick, PhD Thesis, MIT, 1982). It is clear that we have only just begun to scratch the surface of the detailed structure of sub-regular patterns and their cognitive relevance.

bigram description. Although there are several paths through this network from the beginning state on the left to the double-circled end state on the right, the 'loop' back from state 2 to state 1 along with the loop from state 3 to 1 can generate songs with an arbitrary number of *cde ab* notes, followed by the notes *cde fg*. From there, a song can continue with the notes *ab* back to state 1, and so lead to another arbitrary number of *cde ab* notes, all finally ending in *cde fg*. In fact, the transitions between states are stochastic; that is, the finch can vary its song by choosing to go from state 2 back to state 1 with some likelihood that is measurably different from the likelihood of continuing on to state 3. In any case, formally this means that the notes *cde fg* can appear in the 'middle' of a song, arbitrarily far from either end, bracketed on both sides by some arbitrarily long number of *cde ab* repetitions. Such a note pattern is no longer strictly locally testable because now there can be no fixed-length 'window' that can check whether a note sequence 'passes'. Instead of checking the note sequences directly, one must use the memory of the FSA indirectly to 'wait' after encountering the first of a possibly arbitrarily long sequence of *cde ab*s. The automaton must then stay in this state until the required *cde fg* sequence appears. Such a language pattern remains recognizable by a restricted FSA, but one more powerful than a simple bigram checking machine. Such complexity seems typical. Figure 3 displays more fully a second, more complex Bengalese finch song drawn according to the same transition network methodology, this time explicitly showing the probability that one state follows another via the numbers on the links between

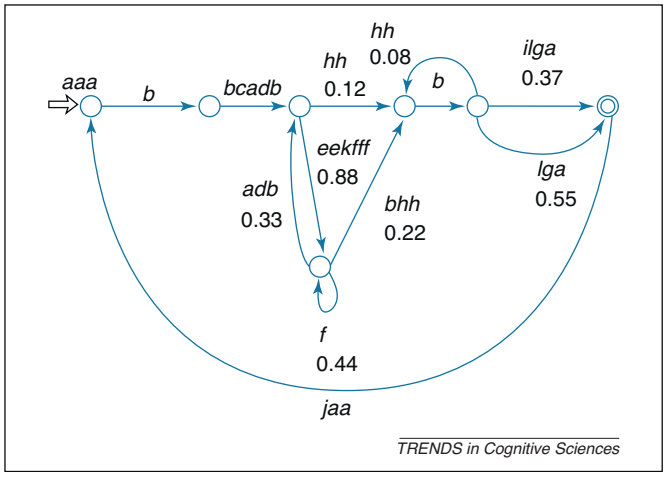

Figure 3. Probabilistic finite-state transition diagram of the song repertoire of a Bengalese finch. Directed transition links between states are labelled with note sequences along with the probability of moving along that particular link and producing the associated note sequence. The possibility of loops on either side of fixed note sequences such as *hh* or *lga* mean that this song is not strictly locally testable (see Box 3 and main text). However, it is still *k*-reversible, and so easily learned from example songs [35]. Adapted, with permission, from [75].

states [32]. It too contains loops, including one from the final, double-circled state back to the start, so that a certain song portion can be found located arbitrarily far in the middle. For example, among several other possibilities, the note sequence *lga,* which occurs on the transition to the double-circled final state, can be preceded by any number of *b hh* repetitions, as well as followed by *jaa b bcadb* and then an arbitrary number of *eekfff adb* notes, again via a loop.

Nightingales, another species with complex songs, can sing motifs with notes that are similarly embedded within looped note chunks [33]. Considering that there are hundreds of such motifs in a song repertoire of a nightingale, their songs must be at least as complex as those of Bengalese finches, at least from this formal standpoint.

More precisely and importantly, the languages involved here, at least in the case of Bengalese finch, and perhaps other avian species, are closely related to constraints on regular languages that enable them to be easily learned [31,34,35]. Kakishita *et al.* [29] constructed a particular kind of restricted FSA generating the observed sequences (a k-reversible FSA). Intuitively, in this restricted case, a learner can determine whether two states are equivalent by examining only the local note sequences that can follow from any two states, determining whether the two states should be considered equivalent [36,37] (Figure I, Box 1). It is this local property that enables a learner to learn correctly and efficiently the proper automaton corresponding to external song sequences simply by listening to them, something that is impossible in general for FSA [38,39].

What about human language sound structure or its phonology? This is also now known to be describable purely in terms of FSA [40], a result that was not anticipated by earlier work in the field [41] which assumed more general computational devices well beyond the power of FSA (Box 1). For example, there are familiar 'phonotactic' constraints in every language, such that English speakers know that a form such as *ptak* could not be a possible English word, but *plast* might be [42]. To be sure, such constraints are often not 'all or none' but might depend on the statistical frequency of word subparts. Such gradation might also be present in birdsong, as reflected by the probabilistic transitions between states, as shown in Figure I (Box 1) and Figure 3 [31,43]. Once stochastic gradation is modelled, phonotactic constraints more closely mirror those found in birdsong finite-state descriptions. Such formal findings have been buttressed by recent experiments with both human infants and Bengalese finches, confirming that adjacent acoustic dependencies of this sort are readily learnable from an early age using statistical and prosodic cues [32,44–46].

However, other human sound structure rules apparently go beyond this simplest kind of strictly local description, although remaining finite state. These include the rules that account for 'vowel harmony' in languages such as Turkish, where, for example, the properties of the vowel *u* in the word *pul*, 'stamp', are 'propagated' through to all its endings [7], and stress patterns (J. Heinz, PhD thesis, University of California at Los Angeles, 2007). Whereas the limited-depth hierarchies that arise in songbird syntax seem reminiscent of the bounded rhythmic structures or 'beat patterns' found in human speech or music, it remains an open question whether birdsong metrical structure is amenable to the formal analysis of musical meter, or even how stress is perceived in birds as opposed to humans [47–49] (Box 4).

Tweets to phrases: the role of words

Turning to syntactic description that lies beyond sound structure, we find that birdsong and human language syntax sharply diverge. In human syntax, but not birdsong, hierarchical combinations of effectively arbitrary depth can be assembled by combining words and words parts, such as the addition of *s* to the end of *apple* to yield *apples,* a word-construction process called 'morphology'. Human syntax then goes even further, organizing words into higher-order phrases and entire sentences. None of these additional levels appear to be found in birdsong. This reinforces Marler's long-standing view [28] that birdsong might best be regarded as 'phonological syntax', a formal language; that is, a set of units (here acoustic elements) that are arranged in particular ways but not others according to a definable rule set.

What accounts for this difference between birdsong and language? First, birdsong lacks semantics and words in the human sense, because song elements are not combined to yield novel 'meanings'. Instead, birdsong can convey only a limited set of intentions, as a graded, holistic communication system to attract mates or deter rivals and defend territory. In terms of the tripartite diagram of Figure 2, the conceptual–intentional component is greatly reduced. Birds might still have some internalized conceptual–intentional system, but for whatever reason it is not connected to a syntactic and externalization component. By contrast, human syntax is intimately wedded to our conceptual system, involving words in both their syntactic and semantic aspects, so that, for example, combining 'red' with 'apples' yields a meaning quite distinct from, for example, 'green apples'. It seems plausible that this single distinction drives fundamental differences between birdsong and human syntax. In particular, birds such as Bengalese finches and nightingales can and do vary their songs in the acoustic domain, rearranging existing 'chunks' to produce hundreds of distinct song types that might serve to identify individual birds and their degree of sexual arousal, as well as local 'dialect-based' congener groups [50–52], although a recent systematic study of song recombination suggests that birds rarely introduce improvised song notes or sequences [32]. For example, skylarks mark individual identity by particular song notes [51], as starlings do with song sequences [52]; and canaries use special 'sexy syllables' to strengthen the effect of mate attraction [50]. However, more importantly, this bounded acoustic creativity pales in comparison with the seemingly limitless open-ended variation observed in even a single human speaker, where variation might be found not only at the acoustic level in how a word is spoken, but also in how words are combined into larger structures with distinct meanings, what could be called 'compositional creativity'. It is this latter aspect that appears absent in birdsong. Song variants do not result in distinct 'meanings' with completely new semantics, but serve only to modify the entirety of the

original behavioural function of the song within the context of mating, never producing a new behavioural context, and so remaining part of a graded communication system. For example, the 'sexy syllable' conveys the strength of the motivation of a canary, but does not change the meaning of its song [50]. In this sense, birdsong creativity lies along a finite, acoustic combinational dimension, never at the level of human compositional creativity.

Second, unlike birdsong, human language sentences are potentially unbounded in length and structure, limited only by extraneous factors, such as short-term memory or lung capacity [53]. Here too words are important. The combination of the Verb *ate* and the Noun *apples* yields the combination *ate apples* that has the properties of a Verb rather than a Noun. This effectively 'names' the combination as a new piece of hierarchical structure, phrase, with the label *ate*, dubbed the *head* of the phrase [54]. This new Verb-like combination can then act as a single object and enter into further syntactic combinations. For example, *Allison* and *ate apples* can combine to form *Allison ate apples*, again taking *ate* as the head. Phrases can recombine *ad infinitum* to form ever-longer sentences, so exhibiting the open-ended novelty that von Humboldt famously called 'the infinite use of finite means' [55], that is immediately recognized as the hallmark of human language: *Pat recalled that Charlie said that Moira thought that Allison ate apples*. Thus in general, sentences can be embedded within other sentences, recursively, as in *the starling the cats want was tired*, in a 'nested' dependency pattern, where we find one 'top-level' sentence, *the starling was tired*, consisting of a Subject, *the starling*, and a Predicate phrase *was tired*, that in turn itself contains a Sentence, *the cats want* formed out of another Subject, *the cats*, and a Predicate, *want*. Informally, we call such embeddings 'recursive', and the resulting languages 'context-free languages' (CFLs; Box 1). This pattern reveals a characteristic possibility for human language, a 'nested dependency'. The singular number feature associated with the Subject, *the starling*, must match up with the singular number feature associated with top-level Verb form *was*, whereas the embedded sentence, *the cats want* has a plural Subject, *the cats,* that must agree with the plural Verb form *want*. Such 'serial nested dependencies' in the abstract form, $a^1a^2b^2b^1$ are both produced and recognized quite generally in human language [53].

The evidence for a corresponding ability in birds remains weak, despite recent experiments on training starlings to recognize such patterns (which must be carefully distinguished from the ability to produce such sequences in a naturalistic setting, as described in the previous section) [56,57]. In starlings, only the ability to recognize nesting was tested, and not the crucial dependency aspect that pairs up particular *a*s with particular *b*s [11] (Box 2). In fact, human syntax goes beyond this kind of recursion to encompass certain strictly mildly context-sensitive constructions that have even more complex, overlapping dependency patterns (Box 1). Importantly, even though they differ on much else, since approximately 1970 a broad range of syntactic theories, comprising most of the major strands of modern linguistic thought, have incorporated Bloomfield's [54] central insight that human language syntax is combinatorially word-centric in the manner described above [13–16,58,59], as well as having the power to describe both nested and overlapped dependencies. To our knowledge, such mild-context sensitivity has never been demonstrated, or even tested, in any non-human species.

In short, word-driven structure building seems totally absent in songbird syntax, and this limits its potential hierarchical complexity. Birdsong motifs lack word-centric 'heads' and so cannot be individuated via some internal labelling mechanism to participate in the construction of arbitrary-depth structures. Whereas a starling song might consist of a sequence of *warbles* and *rattle* motif classes [57], there seems to be no corresponding way in which the acoustical features of the *warble* class are then used to 'name' distinctively the *warble-rattle* sequence as a whole, so that this combination can then be manipulated as single unit phrases into ever-more complex syntactic structures.

Birdsong phrase structure?

Nonetheless, recent findings suggest that birds have a limited ability to construct phrases, at least in the acoustic domain, as noted above, accounting for individual variation within species [32,33]. In particular, there might be acoustic segmentation chunking in the self-produced song of the Bengalese finch [29,31]. Suge and Okanoya used the 'click' protocol pioneered by Fodor *et al.* [60] to probe the 'psychological reality' of syntactic phrases in humans [34].

Box 4. Questions for future research

- We do not know for certain the descriptive complexity of birdsong. Does it belong to any particular member of the sub-regular language hierarchies, or does it lie outside these, possibly in the family of strictly CFLs? If birdsong is contained in some sub-regular hierarchy, how is this result to be reconciled with the findings in the Gentner *et al.* starling study [56]? If birdsong is context free, then we can again ask to what family of CFLs it belongs: is it a deterministic CFL (as opposed to a general CFL)? Is it learnable from positive examples?
- Current tests of finite-state versus CFL abilities in birdsong have chosen only the weakest (computationally and descriptively simplest) finite-state language to compare against the simplest CFL. Can starlings be trained to recognize descriptively more complex finite-state patterns; for example, a locally testable but not non-strictly local testable finite-state pattern, such as $a^1(ba^1)^1$, where a bird would have to recognize a note(s) such as *b* arbitrarily far from both ends of a song [68]? What about sub-regular patterns that are more complicated than this?
- The Gentner *et al.* experiment [49] did not test for the nested dependency structure characteristic of embedded sentences in human language. Can birds be trained to recognize truly nested dependencies, even if just of finite depth?
- Using the methods developed in, for example, [71], what is the descriptive complexity of prosody or rhythmic stress patterns in birdsong?
- What are the neural mechanisms underlying variable song sequences in songbirds? Both human speech and birdsong involve sequentially arranged vocalizations. Are there similar neural mechanisms for the production and perception of such sequences in songbirds and humans? Bolhuis *et al.* [9] have summarized current knowledge of these mechanisms in humans and birds.

Applied to human language, subjects given 'click' stimuli in the middle of phrases such as *ate the apples*, tend to 'migrate' their perception of where the click occurs to the beginning or end of the phrase. Suge and Okanoya established that 3-4 note sequences, such as the *cde* in Figure I (Box 1) are perceived as unitary 'chunks' so that the finches tended to respond as if the click was at the *c* or *e* end of an *cde* 'chunk [34]. Importantly, recall that Bengalese finches are also able to produce such sequence chunks, as described earlier and in Figure I (Box 1) and Figure 3. This is strikingly similar to the human syntactic capacity to 'remember' an entire sequence encapsulated as a single phrase or a 'state' of an automaton, and to reuse that encapsulation elsewhere, just as human syntax reuses Noun Phrases and Verb Phrases. However, Bengalese finches do not seem to be able to manipulate chunks with the full flexibility of dependent nesting found in human syntax. One might speculate that, with the addition of words, humans acquired the ability to label and 'hold in memory' in separate locations distinct phrases such as *Allison ate apples* and *Moira thought otherwise*, parallel to the ability to label and separately store in memory the words *ate* and *thought*. Once words infiltrated the basic pre-existing syntactic machinery, the combinatory possibilities became open ended.

Conclusions and perspectives

Despite considerable linguistic interest in birdsong, few studies have applied formal syntactic methods to its structure. Those that do exist suggest that birdsong syntax lies well beyond the power of bigram descriptions, but is at most only as powerful as *k*-reversible regular languages, lacking the nested dependencies that are characteristic of human syntax [11,29,56,57]. This is probably because of the lack of semantics in birdsong, because song sequence changes typically alter message strength but not message type. This would imply that birdsong might best serve as an animal model to study learning and neural control of human speech [9], rather than internal syntax or semantics *per se*. Furthermore, comparing the structure of human speech and birdsong can be a useful tool for the study of the evolution of brain and behaviour (Box 4). Bolhuis *et al.* [9] have argued that, in the evolution of vocal learning, both common descent (homologous brain regions) and evolutionary convergence (distant taxa exhibiting functionally similar auditory-vocal learning) have a role.

References

1 Darwin, C. (1882) *The Descent of Man and Selection in Relation to Sex*, Murray
2 Margoliash, D. and Nusbaum, H.C. (2009) Language: the perspective from organismal biology. *Trends Cogn. Sci.* 13, 505–510
3 Hauser, M.D. *et al.* (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579
4 Bolhuis, J.J. and Wynne, C.D.L. (2009) Can evolution explain how minds work? *Nature* 458, 832–833
5 Doupe, A.J. and Kuhl, P.K. (1999) Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631
6 Bolhuis, J.J. and Gahr, M. (2006) Neural mechanisms of birdsong memory. *Nature Rev. Neurosci.* 7, 347–357
7 Yip, M. (2006) The search for phonology in other species. *Trends Cogn. Sci.* 10, 442–446
8 Okanoya, K. (2007) Language evolution and an emergent property. *Curr. Op. Neurobiol.* 17, 271–276
9 Bolhuis, J.J. *et al.* (2010) Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Rev. Neurosci.* 11, 747–759
10 Chomsky, C. (1966) *Cartesian Linguistics*, Harper & Row
11 Corballis, M.C. (2007) Recursion, language, and starlings. *Cogn. Sci.* 31, 697–704
12 Aristotle (1970) *Historia Animalium. v.II*, Harvard University Press
13 Steedman, M. (2001) *The Syntactic Process*, MIT Press
14 Kaplan, R. and Bresnan, J. (1982) Lexical-functional grammar: a formal system for grammatical relations. In *The Mental Representation of Grammatical Relations* (Bresnan, J., ed.), pp. 173–281, Cambridge, MA, MIT Press
15 Gazdar, G. *et al.* (1985) *Generalized Phrase-structure Grammar*, Harvard University Press
16 Pollard, C. and Sag, I. (1994) *Head-driven Phrase Structure Grammar*, University of Chicago Press
17 Culicover, P. and Jackendoff, R. (2005) *Simpler Syntax*, Oxford University Press
18 Goldberg, A. (2006) *Constructions at Work: The Nature of Generalization in Language*, Oxford University Press
19 Marr, D. (1982) *Vision*, W.H. Freeman & Co
20 Rogers, J. *et al.* (2010) On languages piecewise testable in the strict sense. In *Proceedings of the 11th Meeting of the Mathematics of Language Association* (eds), pp. 255–265, Springer-Verlag
21 Okanoya, K. (2004) The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Ann. NY Acad. Sci.* 1016, 724–735
22 Sasahara, K. and Ikegami, T. (2007) Evolution of birdsong syntax by interjection communication. *Artif. Life* 13, 259–277
23 Catchpole, C.K. and Slater, P.J.B. (2008) *Bird Song: Biological Themes and Variations*, (2nd edn), Cambridge University Press
24 Wohlgemuth, M.J. *et al.* (2010) Linked control of syllable sequence and phonology in birdsong. *J. Neurosci.* 29, 12936–12949
25 Todt, D. and Hultsch, H. (1996) Acquisition and performance of repertoires: ways of coping with diversity and versatility. In *Ecology and Evolution of Communication* (Kroodsma, D.E. and Miller, E.H., eds), pp. 79–96, Cornell University Press
26 Gentner, T. and Hulse, S. (1998) Perceptual mechanisms for individual vocal recognition in European starlings. *Sturnus vulgaris. Anim. Behav.* 56, 579–594
27 Dobson, C.W. and Lemon, R.E. (1979) Markov sequences in songs of American thrushes. *Behaviour* 68, 86–105
28 Marler, P. (1977) The structure of animal communication sounds. In *Recognition of Complex Acoustic Signals: Report of the Dahlem Workshop on Recognition of Complex Acoustic Signals, Berlin* (Bullock, T.H., ed.), pp. 17–35, Abakon-Verlagsgesellschaft
29 Kakishita, Y. *et al.* (2009) Ethological data mining: an automata-based approach to extract behavioural units and rules. *Data Min. Knowl. Disc.* 18, 446–471
30 Hilliard, A.T. and White, S.A. (2009) Possible precursors of syntactic components in other species. In *Biological Foundations and Origin of Syntax* (Bickerton, D. and Szathmáry, E., eds), pp. 161–184, MIT Press
31 Okanoya, K. (2004) Song syntax in Bengalese finches: proximate and ultimate analyses. *Adv. Stud. Behav.* 34, 297–345
32 Takahasi, M. *et al.* (2010) Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata var. domestica*). *Ethology* 116, 481–489
33 Todt, D. and Hultsch, H. (1998) How songbirds deal with large amount of serial information: retrieval rules suggest a hierarchical song memory. *Biol. Cybern.* 79, 487–500
34 Suge, R. and Okanoya, K. (2010) Perceptual chunking in the self-produced songs of Bengalese finches (*Lonchuria striata var. domestica*). *Anim. Cog.* 13, 515–523
35 Kakishita, Y. *et al.* (2007) Pattern extraction improves automata-based syntax analysis in songbirds. ACAL 2007. *Lect. Notes in Artif. Intell.* 828, 321–333
36 Kobayashi, S. and Yokomori, T. (1994) Learning concatenations of locally testable languages from positive data. *Algorithmic Learning Theory, Lect. Notes in Comput. Sci.* 872, 407–422
37 Kobayashi, S. and Yokomori, T. (1997) Learning approximately regular languages with reversible languages. *Theor. Comput. Sci.* 174, 251–257
38 Angluin, D. (1982) Inference of reversible languages. *J. ACM* 29, 741–765

39 Berwick, R. and Pilato, S. (1987) Learning syntax by automata induction. *J. Mach. Learning* 3, 9–38
40 Johnson, C.D. (1972) *Formal Aspects of Phonological Description*, Mouton
41 Chomsky, N. and Halle, M. (1968) *The Sound Patterns of English*, Harper & Row
42 Halle, M. (1978) Knowledge unlearned and untaught: what speakers know about the sounds of their language. In *Linguistic Theory and Psychological Reality* (Halle, M. *et al.*, eds), pp. 294–303, MIT Press
43 Pierrehumbert, J. and Nair, R. (1995) Word games and syllable structure. *Lang. Speech* 38, 78–116
44 Kuhl, P. (2008) Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843
45 Newport, E. and Aslin, R. (2004) Learning at a distance. I. Statistical learning of non-adjacent regularities. *Cog. Sci.* 48, 127–162
46 Gervain, J. and Mehler, J. (2010) Speech perception and language acquisition in the first year of life. *Ann. Rev. Psychol.* 61, 191–218
47 Halle, M. and Vergnaud, J-R. (1990) *An Essay on Stress*, MIT Press
48 Lerdahl, F. and Jackendoff, R. (1983) *A Generative Theory of Tonal Music*, MIT Press
49 Fabb, N. and Halle, M. (2008) *A New Theory of Meter in Poetry*, Cambridge University Press
50 Kreutzer, M. *et al.* (1999) Social stimulation modulates the use of the 'A' phrase in male canary songs. *Behaviour* 136, 1325–1334
51 Briefer, E. *et al.* (2009) Response to displaced neighbours in a territorial songbird with a large repertoire. *Naturwissenschaften* 96, 1067–1077
52 Knudsen, D.P. and Gentner, T.Q. (2010) Mechanisms of song perception in oscine birds. *Brain Lang.* 115, 59–68
53 Chomsky, N. and Miller, G. (1963) Finitary models of language users. In *Handbook of Mathematical Psychology* (Luce, R. *et al.*, eds), pp. 419–491, Wiley
54 Bloomfield, L. (1933) *Language*, Henry Holt
55 von Humboldt, W. (1836) *Über die Verschiedenheit des menschlichen Sprachbaues* und ihren Einfluss auf die geistige Entwickelung des Menshengeschlechts, Ferdinand Dümmler
56 Gentner, T.Q. *et al.* (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207
57 Gentner, T. (2007) Mechanisms of auditory pattern recognition in songbirds. *Lang. Learn. Devel.* 3, 157–178
58 Chomsky, N. (1970) Remarks on nominalization. In *Readings in English Transformational Grammar* (Jacobs, R.A.P. and Rosenbaum, P., eds), pp. 184–221, Ginn
59 Joshi, A. *et al.* (1991) The convergence of mildly context-sensitive grammar formalisms. In *Foundational Issues in Natural Language Processing* (Sells, P. *et al.*, eds), pp. 31–82, MIT Press
60 Fodor, J. *et al.* (1965) The psychological reality of linguistic segments. *J. Verb. Learn. Verb. Behav.* 4, 414–420
61 Chomsky, N. (1956) Three models for the description of language. *IRE Trans. Info. Theory* 2, 113–124
62 Rogers, J. and Hauser, M. (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 *Recursion and Human Language* (van der Hulst, H., ed.), pp. 213–232, De Gruyter Mouton
63 Huybregts, M.A.C. (1984) The weak adequacy of context-free phrase structure grammar. In *Van Periferie Naar Kern* (de Haan, G.J. *et al.*, eds), pp. 81–99, Foris
64 Shieber, S. (1985) Evidence against the context-freeness of natural language. *Ling. Philos.* 8, 333–343
65 Kudlek, M. *et al.* (2003) Contexts and the concept of mild context-sensitivity. *Ling Phil.* 26, 703–725
66 Berwick, R. and Weinberg, A. (1984) *The Grammatical Basis of Linguistic Performance*, MIT Press
67 van Heijningen, C.A.A. *et al.* (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20538–20543
68 Rogers, J. and Pullum, G. Aural pattern recognition experiments and the subregular hierarchy. *J. Logic, Lang. & Info* (in press)
69 McNaughton, R. and Papert, S. (1971) *Counter-free Automata*, MIT Press
70 Trahtman, A. (2004) Reducing the time complexity of testing for local threshold testability. *Theor. Comp. Sci.* 328, 151–160
71 Heinz, J. (2009) On the role of locality in learning stress patterns. *Phonology* 26, 305–351
72 Crespi-Reghizzi, S. (1978) Non-counting context-free languages. *J. ACM* 4, 571–580
73 Crespi-Reghizzi, S. (1971) Reduction of enumeration in grammar acquisition. In *Proceedings of the 2nd International Joint Conference on Artificial Intelligence* (Cooper, D.C., ed.), pp. 546–552, William Kaufman
74 Crespi-Reghizzi, S. and Braitenburg, V. (2003) Towards a brain compatible theory of language based on local testability. In *Grammars and Automata for String Processing: from Mathematics and Computer Science* (Martin-Vide, C. and Mitrana, V., eds), pp. 17–32, Gordon & Breach
75 Hosino, T. and Okanoya, K. (2000) Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport* 11, 2091–2095