Complex Acoustic Pattern Learning in Songbirds and Humans

Kimberly M. Fenn (kmfenn@uchicago.edu)
Department of Psychology, Center for Cognitive and Social Neuroscience, 5848 S. University Ave.
Chicago, IL 60637 USA

Timothy P. Brawn (tbrawn@uchicago.edu)
Department of Organismal Biology and Anatomy, 1027 E. 57th St.
Chicago, IL 60637 USA

Timothy Q. Gentner (tgentner@ucsd.edu)
Department of Psychology, 9500 Gilman Dr.
La Jolla, CA 92093 USA

Daniel Margoliash (dan@bigbird.uchicago.edu)
Department of Organismal Biology and Anatomy, Center for Cognitive and Social Neuroscience, 1027 E. 57th St.
Chicago, IL 60637 USA

Howard C. Nusbaum (h-nusbaum@uchicago.edu)
Department of Psychology, Center for Cognitive and Social Neuroscience, 5848 S. University Ave.
Chicago, IL 60637 USA

Abstract

European starlings have the ability to learn to classify complex grammatical patterns that follow a context-free grammar (CFG) composed of song motifs (Gentner, et al., 2006) but cotton-top tamarins failed to show similar discrimination when exposed to patterns composed of human speech (Fitch & Hauser, 2004). Beyond differences in species-specific cognitive ability, methodological differences might account for the difference in findings. Here we show that European starlings can learn context-free patterns composed of human speech syllables. We further show that when humans are exposed to the same patterns composed of starling motifs, knowledge of the elements affects classification. Thus, starlings' grammatical ability does not depend on any species-specific aspect of the patterns. Conversely, humans may not have an innate endowment for complex grammatical processing without the aid of explicit training on the elements that compose the patterns.

Keywords: learning; language; syntax; recursion

Introduction

How we acquire knowledge of linguistic patterns has been a controversial theoretical issue in research on language development. One view is that human language cannot be learned—the complexity of the necessary rules to be learned is greater than could putatively be induced on the basis of actual language input (e.g. Chomsky, 1965; Pinker, 1994). The alternative view (e.g. Sokolov & Snow, 1994) is that linguistic experience provides the experience and feedback that is sufficient to learn even the most complex aspects of linguistic structure.

In part, the syntactic complexity of natural language has served as the basis for the argument against learnability (e.g., Chomsky, 1965). Indeed, recently it has been claimed that the computational capacity to process more complex, recursive structures of natural language is a uniquely human faculty (Fitch et al., 2005; Hauser et al., 2002). A recent study reported that following familiarization trials, cotton-top tamarins are able to discriminate a grammatical pattern that follows a finite state grammar, of the form $(AB)^n$ wherein a pair of elements ('A-B') is repeated a small number of times, but are not able to discriminate a more complex context-free structure of the form A^nB^n, in which a number of 'A' elements is followed by an equal number of 'B' elements. The authors argue that tamarins do not have the computational capacity to learn the more complex structure (Fitch & Hauser, 2004).

In contrast, we demonstrated that European starlings are able to learn to classify context-free patterns of the form A^nB^n and reject nongrammatical strings (Gentner et al., 2006) when the patterns are composed of starling song motifs. It is unlikely that starlings exceed the general computational capacity of tamarins, although as vocal learners they may indeed have specialized computational abilities more akin to humans than tamarins. Differences in experimental design may have contributed to the contrasting results. The starlings were trained in an operant procedure on grammatical and nongrammatical strings whereas the tamarins were simply exposed to positive examples. Also, starlings were trained on patterns composed of starling motifs whereas tamarins were exposed to patterns composed of human speech syllables.

For starlings, motifs may not have any meaning on their own but they are familiar sound categories. If the pattern constituents were less familiar would starlings still be able to learn these patterns? Starlings, as acoustic mimics

(Hindmarsh, 1984), will attend to a wide range of sounds in order to reproduce them. However, if acoustic patterns are composed of elements that have little or no intrinsic meaning, it is entirely possible that starlings will not be able to learn the more computationally challenging syntax (CFG). On the other hand, their more general acoustic sensitivity may make attention to arbitrary signals easier. One way to test this is to assess whether starlings can learn both FSG and CFG patterns with human speech syllables. The failure to learn such patterns would indicate that the nature of the constituent elements of these patterns is critical to the process of learning complex syntax.

This raises a complementary question regarding the human uniqueness of computational capacity to process recursion. Fitch and Hauser (2004) demonstrated that humans can discriminate both FSG and CFG patterns following exposure to samples when those patterns are composed of human speech syllables. However, it is not clear how complex syntactic patterns might be learned when the elements are complex and unfamiliar, for example heterospecific sounds. Thus, we can ask if humans can learn FSG and CFG patterns as easily when they are composed of starling motifs as they can when the patterns are composed of speech syllables.

Experiment 1

Humans can learn FSG and CFG patterns comprised of speech syllables and starlings can learn FSG and CFG patterns comprised of starling motifs. However, tamarins can only master the simpler FSG patterns after exposure. To what extent are the differences in results due to species differences in computational capacity and to what extent are the differences due to experience with the pattern constituents or to differences in experience with the patterns themselves?

Method

Participants

Three adult male European starlings were used for this experiment. Prior to testing, subjects were housed in large mixed-sex flight cages along with $15 - 20$ conspecific birds. Subjects were naive to the training and testing stimuli at the start of behavioral training.

Materials

Sound patterns were created using recorded human consonant-vowel (CV) speech stimuli. Different speakers were used for 'A' and 'B' classes. All 'A' stimuli were recorded by a female speaker and 'B' by a male speaker, differing in pitch. Possible CV stimuli for each class were also distinct (see Fitch and Hauser, 2004 for detailed stimulus description). Patterns followed either a finite-state grammar, $(AB)^n$, or a context-free grammar, $A^n B^n$. Two sets of training stimuli were created; one set followed the FSG pattern and one set followed the PSG. The sixteen training stimuli (8 FSG and 8 CFG) were iterations of n=2, either $(AB)^2$ or $A^2 B^2$. Additional stimuli that did not follow either grammar were also created *(AAAA, ABBA, BAAB, BBBB)*.

Behavioral apparatus

All training and testing was conducted in an operant training apparatus that has been described in detail elsewhere (Gentner et al., 2006; Gentner & Margoliash, 2003). An operant panel with three circular response 'buttons' was mounted inside a sound attenuated chamber. Each response 'button' was a PVC housed opening in the panel fitted with an IR receiver and transmitter that detected when the bird broke the plane of the opening with its beak. A remotely controlled hopper moved the food into the subject's reach beneath the opening. Acoustic stimuli were presented to the subject through a small audio speaker mounted out of the subject's view.

Procedure

Starlings were first familiarized with the apparatus through a shaping procedure (Gentner & Margoliash, 2003). A go-nogo operant procedure was used for training (Gentner et al., 2006); subjects learned to respond to the sequences defined by one grammar (S+ stimuli) and to withhold responses to sequences defined by the other grammar (S-stimuli). Starlings pecked the center port to initiate the presentation of a stimulus. For S+ stimuli, pecks to the centre port elicited a food reward for 3 seconds. For S-stimuli, pecks to the centre port elicited a 10-second lights-out punishment during which new stimuli could not be initiated. Sixteen unique stimuli were used during training; eight followed the FSG and eight followed the PSG. The same stimuli were used for all of the birds, but response contingencies were dependent on training condition. The CFG was the S+ and the FSG was the S- for two of the birds and for the third bird, the FSG was the S+ and the CFG was the S-. Trials proceeded until starlings reached stable asymptotic performance.

After achieving asymptote on the baseline stimuli, starlings were transferred to a new set of 16 stimuli (8 of each grammar). If starlings were simply memorizing the trained exemplars, we would not expect to see transfer to novel exemplars. However, if they were learning pattern information, then learning should transfer to new exemplars of the same patterns.

The probe session was designed to test for alternate strategies to solve the task. During training, all stimuli that did not follow the trained grammar necessarily followed the other grammar and the birds could have correctly solved the problem based on a primacy or recency strategy, or by detecting a transition between B/A elements. For example, when trained on the FSG birds could simply respond based on the first two or last two elements (both AB) as all other stimuli began with AA or ended with BB. The transition between elements could have likewise been used to solve the task. A single B/A transition would be sufficient for classification as the FSG contained one such transition but

the CFG contained none. The following four probe stimuli were used to test these strategies: *AAAA, ABBA, BAAB, BBBB*. In addition, a new set of novel grammatical probe stimuli were also used (8 novel stimuli of each grammar).

Results and Discussion

Our initial question was whether starlings would be able to learn the grammatical structures when the elements of the grammars were human speech stimuli. We measured performance using d', (Macmillan & Creelman, 1991). Learning was considered significant when the starlings achieved 5 consecutive blocks (of 100 trials) with $d' > 1.0$ and a lower bound of the 95% confidence interval above zero. Each bird achieved criterion performance, although the number of trials to achieve criterion varied by subject (239 and 311 blocks for the CFG birds, and 203 blocks for the FSG bird). This rate of learning is comparable to that found when starlings are trained on the grammars using starling motifs (Gentner et al., 2006).

We were not interested in whether the starlings could simply memorize the 8 patterns used during training, but whether they could learn the grammatical structure. To assess immediate transfer to novel pattern exemplars, we looked at performance on the first block with the new stimuli. All three birds showed immediate transfer. Average d' during this block was 2.07 (lower CI: 1.19) and 1.48 (lower CI: 0.83) for the birds trained on the CFG and 2.65 (lower CI: 1.77) for the bird trained on the FSG. This is comparable to the birds' performance on the final block of training: 1.46 (lower CI: 0.57), 1.62 (lower CI: 0.97), and 3.07 (lower CI: 0.61), respectively.

The probe session investigated alternate strategies that the birds may have used to perform reliably in this task. Four agrammatical probe stimuli (*AAAA, ABBA, BAAB,* and *BBBB)* were used to test for primacy, recency or B/A transition strategies. Response rates to the agrammatical probe stimuli suggest that the birds clearly discriminate between the agrammatical stimuli and novel S+ stimuli; individual analyses demonstrate that response rates were higher to novel S+ stimuli than to agrammatical stimuli (χ^2 (1) > 11.8, p < .001 for all birds).

We also tested each alternate strategy individually (Figure 1). The two birds trained on the CFG do not show any evidence of using a primacy strategy. These birds had an average d' of 1.2 to the novel n=2 stimuli and only 0.58 to the primacy probe stimuli. However, it appears that the starling trained on the FSG may have been using a primacy strategy. This bird had a d' of 2.35 to the primacy probe stimuli and only 1.05 to novel n=2 stimuli.

Analysis of the recency stimuli suggests that the birds clearly do not treat these stimuli as grammatical. Average d' to recency stimuli is -.673 and all three birds showed negative d' to these stimuli. It is also unlikely that the birds are distinguishing between the grammars by detecting a transition from B to A elements. Average d' to these stimuli was only 0.25 (± .14).
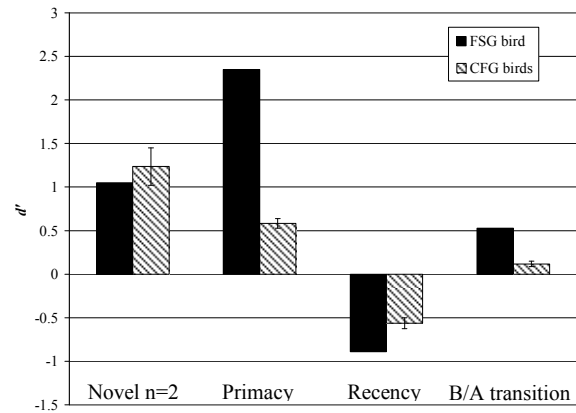

Figure 1: Performance on tests of alternate strategies for the three birds.

Overall, these data provide strong evidence that starlings are able to learn syntactic patterns when the elements of the pattern are human speech syllables. Importantly, we have shown that the birds are not using alternate strategies to solve the more difficult CFG. This provides converging evidence that starlings have the computational capacity to process a complex grammatical structure. However, the change in elements may reflect an increase in classification difficulty for the birds. One of the three birds appears to be using a primacy classification strategy at least some of the time. This result is different than was observed with birds trained on conspecific motifs, suggesting that the human syllables may make learning somewhat more difficult. Because this bird was trained on the simpler FSG, and the two CFG-trained birds did learn the both the CFG and FSG patterns (as reflected in their discrimination of these patterns), we do not believe that the one FSG-trained bird's reliance on a primacy strategy reflects any fundamental limitation in starlings' ability to learn patterns composed of unfamiliar elements, but likely reflects individual variation. Because of the small sample size, we cannot make any strong claims regarding this and future work should address this issue.

Experiment 2

We have shown that European starlings are capable of learning a context free grammatical pattern composed of elements of human speech. Thus starlings and humans (Fitch & Hauser, 2004) have the computational capacity for processing relatively complex syntactic patterns. Is this capacity sufficient to predict pattern classification ability for patterns constructed from complex, unknown acoustic elements? What is the role of perceptual sensitivity to the pattern elements in this capacity? Although humans are capable of classifying a broad range of sound patterns, experience is important in determining those categories. .If humans are presented with patterns composed of starling motifs, can they also learn these grammatical structures? Given that humans possess the computational capacity to process recursive grammatical patterns, they should easily learn the patterns. However, if knowledge of the elements

is important for complex pattern learning, than the ability to learn the patterns may depend on pattern complexity.

Method

Participants

48 students or employees from The University of Chicago volunteered for this study. All participants were native speakers of English and reported no speech or hearing disorders.

Materials

Sound patterns were created using recordings of an adult male European starling. Starling song is composed of four spectro-temporally distinct categories: whistles, warbles, rattles, and high-frequency motifs (Eens, 1997). Eight individual 'rattles' and eight 'warbles' from one starling were used in the patterns. Patterns followed either a finite-state grammar $(AB)^n$ or context-free grammar A^nB^n, where A refers to a rattle motif and B refers to a warble. All stimuli contained only two iterations, either $(AB)^2$ or A^2B^2. Detailed description of stimulus creation has been previously reported (Gentner et al., 2006).

For each grammar, 60 unique sequences were created. The sequences used during training were not used during testing so the expression of learning involved generalization to new stimuli. Furthermore, the 16 individual rattle and warble motifs that comprised the patterns were also used for training and testing separate from the grammars.

Procedure

Participants were randomly assigned to either a motif-training or motif-familiarization group. Pilot testing has shown that without any motif experience, simple exposure to the patterns does not produce learning. To investigate the role of knowledge of constituent elements in pattern classification, one group received explicit training on the motifs. During training, participants first heard one example of a rattle and one example of a warble. They then listened to individual motifs and classified each as being either a rattle or a warble. After responding, feedback was given in the form of a second auditory presentation, accompanied by category information. The second group was only familiarized with the motifs. Participants listened to individual motifs and were told that the sounds would be of two different types, but no feedback was given. For both groups, each warble (n=8) and each rattle (n=8) was presented 3 times, for a total of 48 motif trials. This was followed by a test on the motifs. Participants in the training group were asked to classify each sound as being either a warble or rattle and participants in the familiarization group were asked to classify the motifs into two categories.

Following the motif procedures, both groups were familiarized with one of the two grammatical structures. Thirty patterns were individually presented that followed either the FSG or the CFG syntax. Participants were told

that the sequences followed the same pattern and were instructed to try to determine the nature of the pattern, but were not given explicit information regarding the pattern. After the familiarization phase, they were tested on their knowledge of the grammar. During the test, 60 patterns were played; half of these followed the FSG and the other half followed the CFG. Participants were asked to classify each sound as following the same pattern as the sounds they heard during familiarization or following a different pattern. All stimuli were unique to those encountered during familiarization and no feedback was given.

Results and Discussion

Both groups performed reliably on the motif test (mean $d' = 3.2 \pm 0.11$ and 1.8 ± 0.18 for training and familiarization, respectively), but the group that received training on the motifs performed significantly better than the group that was only exposed to the motifs ($t_{46} = 6.6$, $p < 0.0001$). This suggests that reliable motif discrimination is possible without training but that training produced significant learning beyond this. Therefore, any differences in performance on the grammars may be explained by different abilities in motif classification.

Data from the grammar test was analyzed as a 2-factor ANOVA with motif experience and grammar type as between subjects factors. As can be seen in Figure 2, the group that received training on the motifs performed better than the group that was only exposed to the motifs ($F_{1,44} = 14.7$, $p < .001$). Moreover, without training on the motifs, grammar performance did not exceed zero for either the FSG ($t_{11} = 1.1$, $p > 0.28$) or the CFG ($t_{11} = 0.5$, $p > 0.6$). Only 1 participant in the motif-exposure group achieved d' greater than 1 on this test, whereas in the training group, half of the participants (n=12) performed at this level.
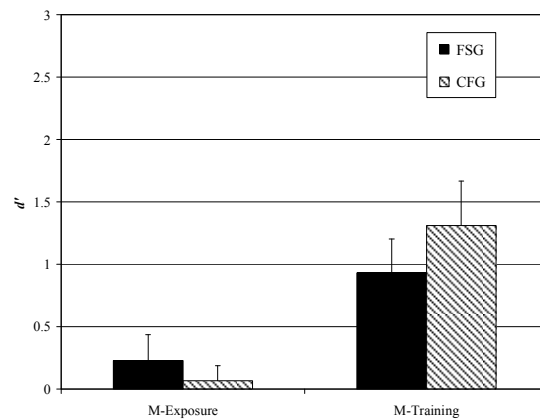

Figure 2: Performance on the grammar test for the motif exposure and motif training groups.

In addition to looking at how performance varied by motif experience, we were also examined whether the amount of learning would vary by syntax. Classification performance did not differ significantly based on grammar ($F_{1,44} = 0.18$, p

> .6). Given that humans can learn both the FSG and CFG reliably well when the elements are human speech and therefore known elements, it may not be surprising that there is no difference in performance after motif training as the elements would be more informative than after motif exposure. However, we might expect performance to be better on the FSG than the CFG when the elements are not known. It is possible that because performance did not exceed chance for either grammar in this group that we are simply seeing floor effects and are unable to detect differences in performance based on grammar.

These results suggest that knowledge of the elements that constitute a pattern significantly affects ability to classify the pattern after exposure. Grammar classification performance was better when it followed training on the starling motifs then when it followed mere exposure to the motifs, even though participants were able to discriminate between the motifs. Even though participants could reliably discriminate between the sounds after familiarization, the motifs were still relatively meaningless, compared to the trained group. This increases the salience of our results; participants are able to perceptually distinguish between the elements, but are still unable to learn the patterns. Explicit knowledge of the elements increases the ability to learn the pattern structure.

When the elements of the pattern have been trained, both the FSG and CFG can be competently learned but when the elements remain relatively meaningless, neither pattern is learned. This may help to explain recent evidence regarding the failure of cotton-top tamarins to learn the CFG pattern used in the present experiment (Fitch & Hauser, 2004). Although we did not find differences in learning based on grammar, it is possible that the tamarins would have performed better on the CFG if the elements of the patterns were more meaningful to them. Although the present data cannot directly support this assertion, the data do suggest that knowledge of the elements affects learning and this may interact with computational complexity of the pattern.

Experiment 3

We have shown that humans can learn syntactic patterns that follow either an FSG or a CFG when the patterns are composed of starling motifs. However, the humans were not tested for alternate strategies. Here we replicate the motif-trained group in Experiment 2 and use agrammatical strings to test if performance can be explained through strategies of primacy, recency, or B/A transition detection.

Method

Participants

The participants were 24 students from The University of Chicago who were native speakers of English and reported no speech or hearing disorders.

Materials

Sound patterns from Experiment 2 were also used in this experiment. Additional stimuli that did not follow either grammar were also created *(AAAA, ABBA, BAAB, BBBB)*.

Procedure

The motif-training procedure used in Experiment 2 was also used in this experiment, with a few modifications. The number of trials in motif-training and grammar exposure was reduced to 32 and 20. Also, the grammar test contained 80 stimuli, 24 each of grammatical FSG and CFG patterns and 32 agrammatical patterns *(AAAA, ABBA, BAAB, BBBB, n= 8 of each)*.

Results and Discussion

Performance on the motif test was again reliable (3.4 ± 1.3 and 3.4 ± 1.2, FSG and CFG, respectively) and did not differ between groups ($t_{22} = .27$, p=.78). To assess performance after grammar exposure, we treated responses from the trained grammar as hits and responses to patterns that followed the untrained pattern as false alarms. We did not include responses to the agrammatical stimuli in the initial assessment of learning. Performance was reliable (1.9 ± 0.34 and 2.1 ± 0.22, FSG and CFG) and did not differ between the groups ($t_{22} = 0.61$, p=0.5)
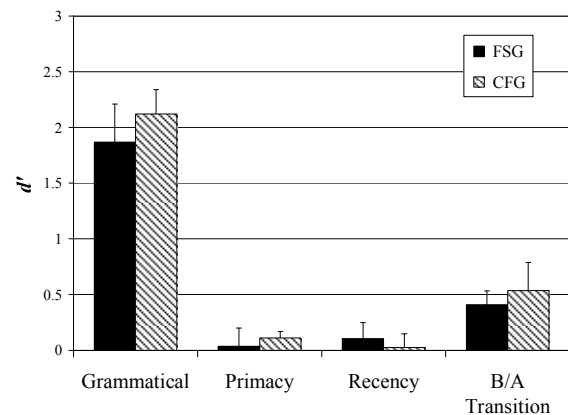

Figure 3: Performance on the grammatical stimuli and agrammatical probe stimuli after motif training and grammar exposure.

We were most interested in assessing whether or not grammar performance could be explained through strategies of primacy, recency or B/A transition detection. Response rates to agrammatical probe stimuli (14.6 ± 4 and 4.7 ± 2.6 percent for FSG-trained and CFG-trained, respectively) were significantly lower than to grammatical stimuli (79.5 ± 5.2 and 82.3 ± 2.6 percent) ($F_{1,22} = 247.5$, p < 0.0001).

Analysis of the individual strategies demonstrates that participants are probably not using alternate strategies in this task (Figure 3). Using a repeated measures ANOVA with stimulus category (grammatical, primacy, recency, B/A transition) as a repeated measure and grammar (FSG or

CFG) as a between subjects factor, there was a significant effect of stimulus category ($F_{1,22} = 65.9$, $p < 0.001$). Post-hoc Sheffe's tests suggest that performance on grammatical stimuli was significantly better than on the primacy, recency, and B/A transition stimuli ($p < 0.0001$ for all comparisons) but performance between the agrammatical stimuli did not differ ($p > .6$ for all). These data strongly suggest that performance on the patterns cannot be explained through the use of alternate strategies.

General Discussion

Vocal patterns are structured sequences of sound elements that are used for communication by songbirds and humans. How is the knowledge of these patterns acquired? The present results, taken together with previous studies, suggest an interaction between the ability to learn a pattern, pattern complexity, and training technique. Operant training produces learning on both simple and complex patterns, despite familiarity with the elements, whereas pattern exposure produces leaning on the more complex syntax only when pattern elements are known. Thus, category information may affect learning in two different ways. Knowledge of pattern elements results in a greater ability to lean patterns from exposure. Without element knowledge, patterns can be learned through operant training. Although operant training does not give explicit category information, reinforcement may help to direct attention to aspects of the signal that are diagnostic for classification.

Linguists have argued that humans cannot learn complex syntax (e.g., Chomsky, 1965; Pinker, 1994) because the complexity of the information to be induced exceeds the information presented in the environment.. The claim that humans have a unique computational specialization for processing recursion (Fitch et al., 2005; Hauser et al., 2002) would seem to be an attempt to make a concrete statement regarding this innate specialization.

The previous demonstration that European starlings share this endowment (Gentner et al., 2006) calls into question the strongest form of this claim regarding general computational capacity for recursion. The present data reject even weaker forms of this claim. First, we have demonstrated that the endowment of recursive capacity does not depend on any species specific aspect of patterns or their elements. European starlings stand ready to learn recursive patterns of human syllables. Their endowment for recursive processing extends beyond a conspecific specialization for bird song. These data demonstrate that starlings' recursive ability goes beyond the ability of a finite-state machine to compute, even if more challenging forms of recursion are yet untested.

The present results also militate against the strongest form of the claim that humans have a broad intellectual endowment for general syntactic processing. While humans have been held up as the paradigm of syntactic processing, there are limits to this ability. When it comes to complex acoustic patterns that are not familiar, humans do not have an innate endowment that can be realized based only on simple exposure to patterns. Furthermore, patterns and their elements are not computationally independent. Syntactic processing may not only require the computational capacity to process the pattern, but the seemingly simple action of processing the elements within the pattern may also be computationally demanding. This may have important implications for research on language acquisition and second language acquisition. Learning the grammatical structure of a language may require additional attentional or memorial demands in processing the sounds or words that compose the grammar.

References

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

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

Chomsky, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, MA: MIT Press.

Eens, M. (1997). Understanding the complex song of the European Starling: an integrated ethological approach. *Advances in the study of Behavior, 26*, 355-434.

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

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

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

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature 424*, 669-674.

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

Hindmarsh, A. (1984). Vocal mimicry in starlings. *Behaviour, 90*, 304-324.

Macmillan, N. A., & Creelman, D. C. (1991). *Detection Theory: A user's guide*. New York: Cambridge University Press.

Pinker, S. (1994). *The Language Instinct*. New York: William Morrow.

Sokolov, J. L., & Snow, C. E. (1994). The changing role of negative evidence in theories of language development. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 38-55). New York: Cambridge University Press.