Acquisition of Phrase Structure in an Artificial Visual Grammar

Sarah T. Wilson (sarahtwilson@gatech.edu)
School of Psychology, 654 Cherry St., J.S. Coon Building, Atlanta, GA 30332
Carla L. Hudson Kam (Carla.HudsonKam@ubc.ca)
Department of Linguistics, 2613 West Mall, Vancouver, BC, V6T 1Z4

Abstract

Recent studies showing learners can induce phrase structure from distributional patterns (Thompson & Newport, 2007; Saffran, 2001) suggest that phrase structure need not be innate. Here, we ask if this learning ability is restricted to language. Specifically, we ask if phrase structure can be induced from non-linguistic visual arrays and further, whether learning is assisted by abstract category information. In an artificial visual grammar paradigm where co-occurrence relationships exist between categories of objects rather than individual items, participants preferred phrase-relevant pairs over frequency-matched non-phrase pairs. Additionally, participants generalized phrasal relationships to novel pairs, but only in the cued condition. Taken together these results show that learners can acquire phrase structure in a non-linguistic system, and that cues improve learning.

Keywords: statistical learning, language learnability, syntax, modality independence

Introduction

Theories of syntax differ, however, most contain two important elements: words are members of categories (traditionally nouns, verbs, determiners, etc.) and these categories are related to each other in higher-order patterns, e.g., phrases or sentences. To give an example in English take the sentence "The cat chased the dog." The word "cat" is a member of the word class, or category, noun and it has a relationship with "the" – its determiner – forming a noun phrase. A similar relationship exists between "the" and "dog." The verb phrase is comprised of "chased" plus "the dog." Thus, the sentence consists of several phrases defined over categories, arranged hierarchically.

In the traditional view, these elements of language are not learned, but rather considered to be innate by necessity (e.g. Crain, 1992; Wexler, 1991). A number of recent studies have begun to challenge the notion that these aspects of language are unlearnable, however, particularly with respect to categories (see, e.g., Mintz, 2002). The other basic properties of syntax, namely phrases (the property of interest in the current study) and their hierarchical organization have proved more challenging for a learning account. Saffran (2001) created a miniature artificial language, based on one used by Morgan, Meier, and Newport (1987), that was defined by a grammar over classes of words. Phrase structure in this language was defined by a number of rewrite rules over a basic or canonical sentence type: S → AP + BP + (CP), where AP, BP, and CP are phrases, and CP is an optional phrase. The phrase rewrite rules were: AP→ A + (D); BP → CP + F or BP → E; and CP → C + (G). Learning of this grammar was statistically above chance; however, it was only marginally so, leaving open the question of whether phrase structure is an innate component of human knowledge.

More recently, Thompson and Newport (2007) used an adapted version of the same language with stronger cues to phrase boundaries – in particular, phrases tended to hang together in perfectly predictive relationships, while various language-like sentential manipulations created dips in predictive dependencies across phrase boundaries that were relatively low – and found greatly enhanced learning.

More specifically, the Thompson and Newport (2007) language had a phrase structure where phrases were composed of pairs of categories of words. There were 6 categories (labeled here, for simplicity: A, B, C, D, E, and F) which formed three phrases: AB, CD, and EF. Categories were distributionally defined. That is, the only way in which words were in the same category was that they occurred in the same locations both absolutely (their place in the sentence) and relatively (their adjacency to other elements). There were a total of 18 monosyllabic words in the language, 3 per category. Phrases could take part in a variety of operations: (1) movement, (2) repetition, (3) omission, and (4) insertion, thereby creating a set of sentences where the probability of a transition between categories within phrases was high (perfect 1.0) and the probability of a transition between categories that occurred across phrase boundaries was low. Importantly, the probability of a transition between individual words was also low, both within and across phrases. Therefore, the only indicator of structure was the transitional probabilities between categories of words — a higher-order relationship. At test, adult participants selected pairs of words which comprised a grammatical phrase more often than pairs of words which had co-occurred equally often in the input but which did not form a phrase, demonstrating they had acquired an understanding of category-level relationships. That is, they had learned categories as well as which categories formed phrases and which did not.

We investigate whether higher-order category relationships of this type are learnable in a non-linguistic system, something that might be expected if such learning is domain general. We exposed participants to visual stimuli constructed to have the same properties as the auditory language used by Thompson and Newport (2007). Simple

two-dimensional objects were organized into categories, then arranged into visual arrays according to a phrase structure grammar based on how categories of objects co-occurred. After exposure, participants were tested to see if they had learned the category-based grammar governing the combination of the items in the array.

We also assessed whether and how learning was affected by the presence and reliability of (non-distributional) cues to category membership. In previous work on larger versions of auditory languages (i.e., languages with a greater number of words per category than Thompson & Newport, 2007) we found that phrase learning is affected by the presence and reliability of cues to category membership (Wilson & Hudson Kam, 2009, 2013). Presumably, the cue makes it easier for people to identify the categories, thereby facilitating the tracking of probabilities over the categories necessary for phrase learning. We were interested in whether this would also be true of learning in the context of a non-linguistic visual system, and so included subtle visual cues to category membership in varying degrees in different conditions.

The visual array paradigm used here is based on that originally developed by Fiser and Aslin (2001). In their third and final experiment, Fiser and Aslin exposed adult participants to a set of visual arrays in which the adjacency relationships had a specific statistical structure irrespective of absolute spatial location. There were 12 uniquely-shaped black objects. Pairs of objects formed base pairs, always appearing together, in one of three possible alignment types: (1) vertical, (2) horizontal, or (3) oblique (diagonal). Additionally, the frequencies of some base pairs and cross-pair, non-base pairs of items were equated. Therefore, the lower order, joint probability of these base pairs and cross pairs were equal (i.e., P(object1, object2) = P(object2, object3)), but the higher-order relative statistic, their conditional probabilities, differed (i.e., P(object2|object1) = 1.0 vs. P(object2|object3) ~ low). At test, participants reliably chose the base pairs over cross pairs, suggesting they understood the higher order conditional probability relationship. (See Figure 1 for a schematic of a sample exposure scene.)

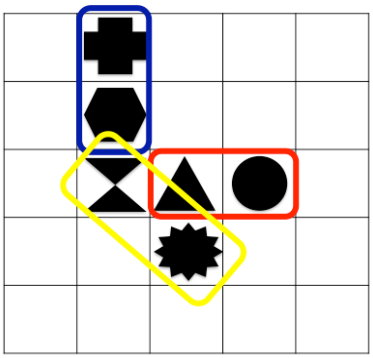

Figure 1. Schematic of example scene from Fiser and Aslin (2001), composed of three base pairs (one vertical, one horizontal, one oblique)

Their paradigm was modified here to investigate the acquisition of a phrase structure, where statistical relationships occur across pairs of categories, as opposed to pairs of individual items. To implement these ideas in the visual array paradigm, we expanded base pair relationships to include categories of objects which were adjacent in relevant configurations, while equating the co-occurrence of individual items within and across phrase boundaries. If our hypothesis is correct, that the learning processes that contribute to learning phrase structure are domain general, then we expect learning outcomes in the visual system to be commensurate with those found in previous auditory artificial language learning work, namely that it is possible to learn from dips in transitional probability that occur between categories of items in order to understand category relatedness (i.e. phrases) and that this learning is facilitated by non-distributional cues to category membership.

Methods

Participants

A total of 60 adults (20 per condition) participated in this study for course credit in Psychology courses at the University of California, Berkeley.

Stimuli

Twenty-four unique objects were used, each with a unique color (properties of the color to be discussed later). Objects were assigned to one of eight categories (A, B, C, D, E, F, G, and H), with three objects per category. Pairs of categories were then grouped into phrases (much like the previous experiments), in one of two forms: vertical or horizontal. Phrases were then arranged into one of 16 distinct arrays in a five by five grid, with each array containing one example of each phrase. The 16 arrays, or category constructions, are much like sentence types. As such, the arrays constitute the 'grammar' of the visual system. Four distinct example arrays are shown in Figure 2.

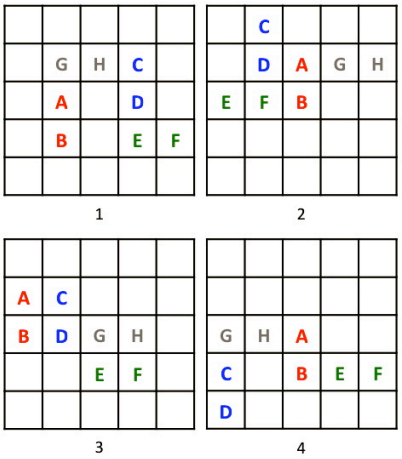

Figure 2. Four examples of the 16 construction types or arrays with category placement labels.

This design resulted in conditional probabilities of adjacent co-occurrence of categories within phrases being perfect (1.0). Adjacent co-occurrence of pairs of categories that were possible but not necessary – i.e, which crossed a phrase boundary - had much lower conditional probabilities: each occurred exactly once over the exposure set, and therefore with p =.0625. The complete set of adjacent co-occurrence relationships, for both the vertical and horizontal dimensions appear below in Tables 1 and 2.

Table 1. Adjacent co-occurrence conditional probabilities, vertical from top category to bottom category (phrase transitions in **bold**)

	A	B	C	D	E	F	G	H
A	-	**1.0**	-	-	-	-	-	-
B	-	-	.06	-	.06	.06	.06	.06
C	-	-	-	**1.0**	-	-	-	-
D	.06	-	-	-	.06	.06	.06	.06
E	.06	-	.06	-	-	-	.06	.06
F	.06	-	.06	-	-	-	.06	.06
G	.06	-	.06	-	.06	.06	-	-
H	.06	-	.06	-	.06	.06	-	-

Table 2. Adjacent co-occurrence conditional probabilities, horizontal from left category to right category (transitions in **bold**)

	A	B	C	D	E	F	G	H
A	-	-	.06	.06	.06	-	.06	-
B	-	-	.06	.06	.06	-	.06	-
C	.06	.06	-	-	.06	-	.06	-
D	.06	.06	-	-	.06	-	.06	-
E	-	-	-	-	-	**1.0**	-	-
F	.06	.06	.06	.06	-	-	.06	-
G	-	-	-	-	-	-	-	**1.0**
H	.06	.06	.06	.06	.06	-	-	-

The adjacent co-occurrence frequencies (or joint probabilities) of some within-phrase pairs and cross-phrase pairs of objects were equated. In order to accomplish this, some object pairs (pairing of particular objects either within or across phrases) were highly frequent (occurring 26 times) and some were less frequent (occurring 6 times). In this way, the less frequent within-category object pairs had equal joint probability as some cross-phrase object pairs (those that occurred adjacently in the 6 examples of any given scene) and served as test items. Additionally, some object pairs, both within phrase and across phrase boundaries, were reserved from the exposure set also for test purposes.

The exposure set contained 96 unique scenes total, 6 of each construction type. (An example scene appears in Figure 3.) The exposure set was seen a total of four times, and so each scene appeared four times per session. All cross-phrase object pairs occurred 24 times per exposure session. Within-phrase object pairs occurred either 24 or 104 times per exposure session. Each individual object occurred exactly 32 times in the exposure set, and so occurred exactly 128 times per exposure session.

Each slide was seen for 2.5 seconds, and was interspersed with 1 second fixation slides. Additionally, there was a 2 minute break at the halfway point. The total exposure session lasted for approximately 25 minutes.

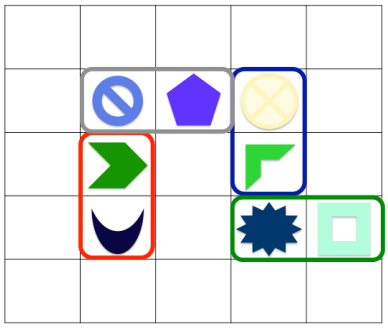

Figure 3. Example visual array (of construction type 1 from Figure 2), with phrases outlined

Note that the visual displays merely appear as complex designs; there is nothing in the visual arrays themselves that indicates the phrasal structure. If anything, Gestalt principles (Palmer, 1999) might lead participants to 'mis-segment' individual arrays into components larger than the phrases. In Figure 2 array 3, for example, participants might perceive two squares rather than four phrases, or in the display in Figure 3 participants might see an archway.

Experimental Manipulation

This study also addressed the contribution of a subtle non-distributional cue to category membership in acquisition of the phrase structure. The visual cue to category was an aspect of the color of the objects irrespective of hue. Colors for objects were selected from levels of brightness and saturation available in Microsoft Powerpoint — three hues from each level. In the cue-present version of the visual arrays, objects from the same category were of different hues from the same brightness and saturation level. In the without cue condition, objects were randomly assigned to categories, therefore, color could not serve as a cue to category membership. A third version of the arrays contained a partially predictive cue to category membership, where two of the three objects in the category were of the same brightness and saturation level.

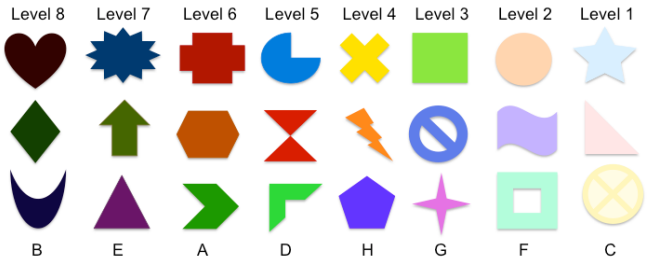

Figure 4. All 24 objects, shown in respective color assignment, organized into 8 levels of brightness and saturation, (category shown at bottom of column).

Tests

There were two types of tests in this experiment designed to test whether participants understood the phrases or units of the visual grammar – very much like the phrase tests from Thompson and Newport (2007). Both tests required participants to compare two pairs of objects: one with a high category-level conditional probability and one with a low category-level conditional probability. The two comparison pairs were displayed to the left and to the right of the center square of the 5 x 5 grid, as shown in Figure 5.

Phrase Test. Some pairs of objects in the exposure set were matched for frequency – that is, had the same joint probabilities of appearing together – either within or across a phrase boundary. However, the pairs differed in that some had high category-level conditional probability (i.e., they were within a phrase) while others had a category-level conditional probability that was low (i.e,, they were not within a phrase). The first test compared these two types of pairs. There were 12 such items total, six on the first day and six on the second day.

Generalization Test. The second test was a generalization test, in which participants were tested using pairs of objects that had been reserved from the exposure set. One test pair was a novel object pair with high category-level conditional probability. The comparison pair of objects was also novel, but with a low category transitional probability (but not zero or absent). There were 12 of these items, six on the first day and six on the second day.

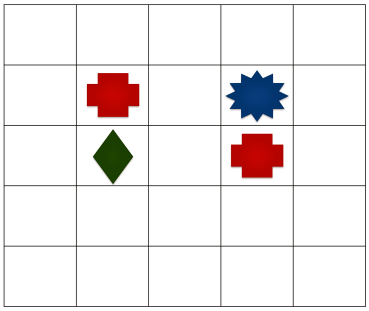

Figure 5. Sample test item, within-phrase object versus frequency matched objects crossing a phrase boundary (vertical phrase).

Procedure

Participation in this study spanned two days, with each day involving an exposure session and a test session. Unlike earlier experiments that tested strictly end-state performance outcomes, we also were interested in the trajectory of learning – whether we could capture an intermediary stage of having learned some aspects, but not all, of the grammar.

On each day, participants saw the exposure set a total of eight times: four times through, followed by a two-minute break, then another four times through, for a total exposure session of about 25 minutes. Across both days, participants saw the exposure set 16 times. After exposure on both days, participants were given the two-alternative, forced choice test.

The phrase test items were always given first, followed by the generalization test items. Prior to test, participants were shown a practice comparison that contained objects that had not appeared in the scenes, first in the vertical then the horizontal orientation. Participants were instructed that they were going to indicate which of the pairs of objects they thought more likely came from the scenes they had been learning about. Responses were recorded by the experimenter, who was also advancing the test-item slides. Participants were given as much time as they needed to make a response.

Results

First, it is of interest to compare performance on the initial phrase test both across the two days and across conditions. This test compared pairs of objects with either high or low category-level conditional probability, with test pairs in the comparison having appeared with the same frequency in the exposure set. Importantly, successful performance on this test cannot be accounted for by simple adjacency since both pairs in the comparison had occurred an equal number of times in the exposure. Mean performance outcomes on this test appear in Figure 6.

An overall, 2 x 3 (day x cue-condition) ANOVA revealed a significant interaction between the two factors in the analysis ($F(5, 119)=3.93$, $p=.022$. (An examination of main effects, day and cue condition, revealed that there were no significant differences ($F(1, 119)=.5452$, $p=.463$ and $F(2,119)=.094$, $p=.910$ respectively). This was also true for simple main effects of condition on both days ($F(2, 59)=1.640$, $p=.203$) and $F(2, 59)=1.936$, $p=.154$).) The interaction reflects the difference in performance patterns for the groups by day, which was, interestingly, not significant for either day. However, given that it was our expectation that all or some of the cue groups would demonstrate learning of the phrases, given results from previous work with auditory languages where this type of distinction was possible, while allowing for differences in performance, we did performance comparisons for each cue group against chance level performance. On Day 1, Without Cue participants performed significantly above chance, M = 63.3%, SD = 48.4% ($t(19) = 2.707$, $p = .014$), while With

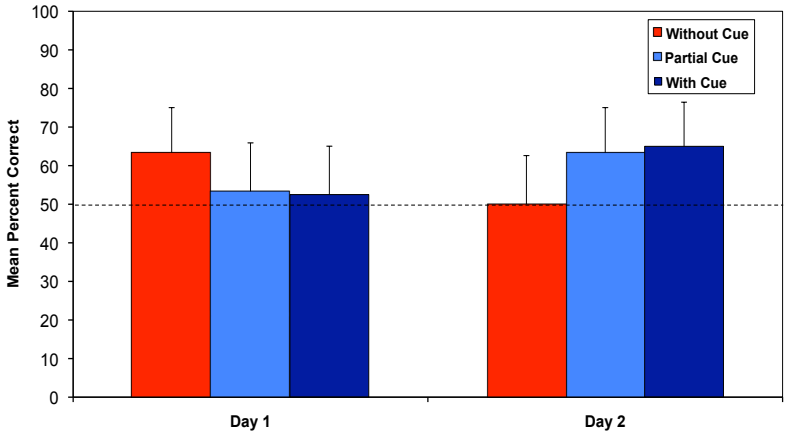

Figure 6. Mean percent correct on the first phrase test. Dashed line indicates chance level performance.

Cue participants performed at chance level, M = 52.5%, SD= 50.1% (t(19) = .529, p = .603) as did Partially Predictive Cue participants, M = 53.3%, SD = 50.1% (t(19) = .748, p = .464).

These means from the second day were also tested against chance performance. Without cue participants performed at chance level, M = 50.0%, SD = 50.2% (t(19) = .000, p = 1.000), while With Cue participants performed above chance, M = 65.0%, SD = 47.9% (t(19) = 2.932, p = .009) as did Partially Predictive Cue participants, M = 63.3%, SD = 48.4% (t(19) = 2.320, p = .032).

We also tested participants' ability to generalize to novel phrases. This test asked participants to compare novel base pairs that had been reserved from the exposure set, but which again differed in that one had a high category-level conditional probability and one had a low category-level conditional probability. Mean performance scores on this test can be seen in Figure 7. An overall, 2 x 3 (day x cue-condition) ANOVA did not reveal a significant interaction (F(5,119)=.173, p=.841). Nor were there main effects of day or cue-condition (F(1,119)=.640, p=.425 and F(2,119)=2.404, p=.095). Simple main effects of condition, additionally, were null for each day (F(2, 59)=1.862, p=.165 and F(2, 59)=.646, p=.528). Nonetheless, there were some intriguing patterns in the data that we pursued further with individual group analysis. As before, we performed planned comparisons to chance. With Cue participants performed significantly above chance on the first day (M = 62.5%, SD = 48.6% (t(19) = 2.380, p = .028)) while Without Cue performed at chance M = 49.2%, SD = 50.2% (t(19) = -.188, p = .853), as did the Partially Predictive Cue participants (M = 51.7%, SD = 50.2% (t(19) = .302, p = .766)).

We also compared performance on the generalization test for the second day. On this day, With Cue, Without Cue, and Partially Predictive Cue participants all scored at chance level (M = 55.8%, SD = 49.9% (t(19) = 1.234, p = .232); M = 48.3%, SD = 50.2% (t(19) = -.302, p = .766); and M = 49.2%, SD = 50.2% (t(19) = -.165, p = .871), respectively.

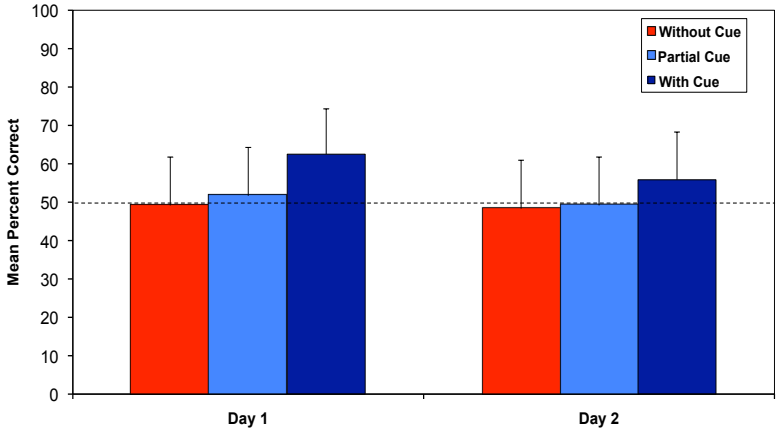

Figure 7. Mean percent correct on the second phrase test

Discussion

This experiment was designed to assess whether category relatedness or phrases can be inferred in a nonlinguistic system, or is instead a property only of linguistic systems. In addition, we asked whether cues to category membership would function similarly in the auditory and visual domains. Participants were exposed to visual arrays comprised of phrases defined over categories, arranged so that the within-phrase category-level conditional probabilities were higher than those of categories that co-occurred but did not form phrases. Participants were then tested to see if they had acquired the phrases of the visual grammar. The hypothesis was that general purpose learning processes would enable the acquisition of phrase structure in the visual system as in the auditory language, and that these learning processes would be improved by cues that facilitated the matching of items in categories. If this is the case, the relative statistics in the input should inform judgments about category relatedness that contrast pairs of objects that are a phrase-relevant pair to pairs that cross phrase boundaries.

We found some evidence of this. On the first day, Without Cue participants performed above chance on the first phrase test, demonstrating that they had learned something about the category-level co-occurrence probabilities. Surprisingly, performance in this group dropped on the second day – potentially the result of looking for further patterns in the stimuli that were not present. In contrast, With Cue and Partially Predictive Cue participants performed at chance level initially on Day 1 and went on to improve on the first phrase test on Day 2. These groups may have taken longer precisely because of the presence of distributional cues that were correlated with the color cue – they were figuring out that relationship first (as demonstrated by their above-chance performance on the second test), then having attained some (albeit shaky) knowledge of the color relationships, they went on to learn the relationships between categories. The With Cue participants were the only group to demonstrate above chance learning on the second test at all, on the first day – a result that may just be due to chance. However, since the relative pattern of performance (With Cue participants doing better) was consistent on this test across the two days we think that the fact that they performed better than the other two groups of participants on this test (even if not significantly so) is a real, if small, effect.

Given that the effects are sometimes present, sometimes absent, it may bring up the question as to whether there were particular aspects of our test stimuli that could have skewed the pattern of the data. There were a number of controls in place to minimize this possibility. While the particular test items were different for all three cue-conditions, the number of pairings that incidentally paired objects of the same hue (albeit different brightness and saturation – as the cue dictated) were the same across all three conditions and all tests and were a very low number. Additionally, each test had two versions: an A version as well as a B version, and those versions were randomized as to whether a particular participant received the A version on Day 1 or the B version. Thus, the pattern of results seems unlikely to have occurred due to particular test stimuli.

Ultimately, these general learning results should be replicated with different participants and stimuli if possible. The explanation for learning being sometimes present, sometimes absent should be explored and tested, possibly by looking at more individual learning trajectories. This project was intended to provide a visual analogue of both our previous work and work by Thompson and Newport (2007) – all of which provided a much longer input period. And so, this work may benefit from equivalent time on task to see if learning improves and generalization ability ever emerges and remains persistent in this paradigm.

Acknowledgments

This work was supported by the National Institutes of Health Grant HD 048572 and a Discovery Grant (Individual) from the Natural Sciences and Engineering Research Council of Canada to CLHK.

References

Crain, S. (1992). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*, 597-650.

Fiser, J. & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*, 499-504.

Mintz, T.H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, *30*, 678-686.

Palmer, S. E. (1999) *Vision science: Photons to phenomenology*. Cambridge, MA: Bradford Books/MIT Press.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493-515.

Thompson, S.P & Newport, E.L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*, 1–42.

Wexler, K. (1991). On the Argument from the Poverty of the Stimulus. In Kasher, A. (Ed.) *The Chomskyan Turn* (252-70), Cambridge: Blackwell Publishers.

Wilson, S. T. & Hudson Kam, C. L. (2013). *Learning Syntax Through Statistics: When do transitional probabilities need a boost?* Unpublished manuscript, Georgia Institute of Technology, Atlanta, GA.

Wilson, S. T. & Hudson Kam, C. L. (2009). *Learning Syntax Through Statistics: When do transitional probabilities need a boost?* Paper presented at the Boston University Conference in Language Development, Boston, MA