1. Introduction

The notion that children use statistical distributions present in the input to acquire various aspects of linguistic knowledge has received considerable recent attention (Saffran, Aslin & Newport 1996; Redington, Chater & Finch 1998; Maye & Gerken 2000; Gomez 2002; Maye, Werker & Gerken 2002; Mintz, Newport & Bever 2002; Mintz 2003; Swingley 2005; among others). What remains unclear is the outcome of learning statistical distributions. At least two possibilities exist. One is that learners use these acquired statistics to create an illusion of structure (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett 1996). A second possibility is that learners use statistics to identify particular abstract syntactic representations (Miller & Chomsky 1963; Yang 2006; Pearl 2007). To investigate this more general question, we focus on how infants acquire the phrase structure (PS) of their target language.

By fourteen-months of age, infants already begin to demonstrate sensitivity to properties of their native language syntax (Hirsch-Pasek & Golinkoff 1996); unfortunately, the nature of the cues that children exploit to infer the structural properties of the target grammar remain poorly understood. Earlier research emphasized the necessary role of semantic (Pinker 1984) and/or prosodic (Gleitman & Wanner 1982; Gleitman, Gleitman, Landau & Wanner 1988; Morgan 1986; Peters 1983) cues in driving the acquisition of phrase structure. More recently, it has been shown that infants can use statistical distributions between word categories to derive the phrase structure of an artificial language grammar (Gomez & Gerken 1999; Thompson & Newport 2007). In a comparison of the utility of distributional and linguistic cues, Morgan, Meier & Newport (1987) found that adults were able to acquire the constituency of artificial languages only when the distributional information was augmented with correlated semantic, prosodic or morphological cues.

Recently, however, Thompson & Newport (2007) suggested that the distributional cues in those experiments were simply not strong enough, in and of themselves, to be informative. Instead, they showed that adult language learners could, indeed, exploit transitional probabilities in acquiring an artificial phrase structure grammar. Our primary criticism of the artificial grammar used in Thompson & Newport (2007) is that it contained phrases with no internal structure, a hallmark of natural language syntax. Therefore, these findings leave
unresolved whether learners can detect statistical cues to internally structured phrases.

In this paper, we first review adult experiments from Takahashi & Lidz (2007), which suggest that adults are able to generalize beyond the input. Then, we present new results from an experiment with 18-month-old infants and a set of computational simulations using a Simple Recurrent Network (SRN). We show that infants can learn nested hierarchical phrase structure by using statistics alone, while the SRN fails. Finally, we conclude by suggesting that the output of learning derives from an interaction between the input and inherent representational system. In other words, infants use these acquired statistical distributions to help them select particular abstract syntactic representations.

2. Artificial Languages

For Takahashi & Lidz (2007) and the experiments reported here, we created two artificial languages (Grammar 1 (G1) and Grammar 2 (G2)) that have an identical canonical sentence (ABCDE) and differed only in constituent structure.

For instance, AB is a constituent in G1 but not in G2. On the other hand, BC is a constituent in G2 and a non-constituent in G1. Importantly, the two grammars show a nested hierarchical structure: CD is a constituent inside a larger constituent CDE in G1, as an example. Both grammars contained the same six lexical categories with three nonsense words each, adapted from Thompson & Newport (2007).

<table>
<thead>
<tr>
<th>Table 1: Nonsense words assigned to each word class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Class</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>KOF</td>
</tr>
<tr>
<td>DAZ</td>
</tr>
<tr>
<td>MER</td>
</tr>
</tbody>
</table>

The sentences contained no prosodic or semantic cues to phrase boundaries. There was an interval of 30 ms between words and an interval of 1400 ms between sentences.
The input included rules that are attested in natural languages. Specifically, constituents could be optional, repeated, substituted by proforms, or moved to the front of the sentence (following Thompson & Newport (2007)). Therefore, the transitional probabilities (TP) between words within phrases were higher than the TPs across phrases. All the experiments presented here use these same two artificial languages throughout.

Table 2: Transitional probabilities for all sentences in G1 and G2

<table>
<thead>
<tr>
<th></th>
<th>A-B</th>
<th>B-C</th>
<th>C-D</th>
<th>D-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>1.00</td>
<td>0.63</td>
<td>1.00</td>
<td>0.71</td>
</tr>
<tr>
<td>G2</td>
<td>0.71</td>
<td>1.00</td>
<td>0.32</td>
<td>1.00</td>
</tr>
</tbody>
</table>

3. Adult Experiments

The first adult experiment asked whether learners can detect statistical cues to nested hierarchical structures in an artificial language grammar (see Takahashi & Lidz (2007) for details). Half of the adult participants heard G1 during the 36-minute familiarization period, while the other half heard G2. The input consisted of an auditory presentation of 80 sentences, repeated six times.

At test, participants were presented with a pair of sentences and asked to decide which sentence was consistent with the familiarized grammar. Here, we focus only on test items in which a constituent versus non-constituent was moved. In other words, at test, one sentence would have a constituent in the familiarization period moved to the front of the sentence, while in the other, a non-constituent in the familiarization period was displaced to the front of the sentence.

![Figure 3: PS tree of a movement sentence in G1](image-url)
Participants (n = 44) were instructed to choose the one that they think belongs to the language they had heard. For example, if you were assigned to hear G1 during the familiarization, you should choose CDEAB, because CDE, a constituent in G1, is moved to the front, whereas if your input grammar was G2, you should choose DEABC because DE, a constituent in G2, is moved to the front. In natural language, movement of a constituent is a possible rule, but movement of a non-constituent is impossible. In this way, one of the test pairs was always the correct answer for one of the grammars and the other was the correct answer for the other grammar. We should also point out that all the word sequences of test items were novel. So the TP between any adjacent test words was always 0.

The results of the adult experiments presented in Takahashi & Lidz (2007) showed that adults correctly picked the consistent answer significantly more often than chance (p < 0.05).

In this experiment, the strings of the test items were novel, but the structure was not. Therefore, we do not know whether learners can project beyond what is in the input, and if so, what generalizations can they conclude. Thus, we
conducted another adult experiment in which the input excluded all sentences that are generated via movement rules. The familiarization did not include any movement-type sentences, but the adults were tested on the same test items as in the previous experiment. In this way, both the structure and strings of the test items were new.

At least two outcomes can be predicted for this experiment. One possibility is if the learning is solely based on what is seen in the input, then both test sentences should be equally illicit since neither is familiar. On the other hand, learners might allow novel structures as long as they are compatible with some inherent constraints on possible linguistic representations. Therefore, at test, participants might choose the test sentences with moved constituents over the sentences with displaced non-constituents. Again, this is because moving a constituent is an attested rule in natural language, while displacement of a non-constituent is impossible in natural languages.

We found that adults (n = 44) chose the test sentence in which a constituent in their familiarized grammar was moved significantly more often than sentences in which a non-constituent was displaced, despite the lack of movement sentences in the input (p < 0.001).

![Figure 6: Results of the second adult experiment](image)

Crucially, knowing the constituency of the artificial language alone cannot give this result. The knowledge that moving a constituent is licit but moving a non-constituent is illicit is also required, suggesting that adults must rely on some type of inherent knowledge to help them decide between the test items and ultimately inform the representations they construct.

4. Infant Experiment

In the infant experiment reported here, we tested to see whether they can learn internally-structured phrase structure on the basis of statistical information alone.
4.1. Participants

Fourteen infants, approximately 18 months of age were tested (range: 17;15–19;09, mean: 18;17). Eight additional infants were tested but excluded from analyses for the following reasons: crying (n = 4), inattentiveness (n = 3) and equipment failure (n = 1). The infants were randomly divided into two groups. Half of the infants heard Grammar 1 as input during the familiarization period and the other half heard Grammar 2.

4.2. Materials

The artificial languages used in the infant experiment were identical to the ones in the adult experiments. In this experiment, the familiarization input included movement sentences. The only difference was that 30 sentences, instead of 80 sentences, were picked as the presentation set. The TP patterns of the presentation set are given in Table 3.

Table 3: Transitional probabilities for 30 input sentences in G1 and G2

<table>
<thead>
<tr>
<th></th>
<th>A-B</th>
<th>B-C</th>
<th>C-D</th>
<th>D-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>1.00</td>
<td>0.22</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>G2</td>
<td>0.25</td>
<td>1.00</td>
<td>0.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The movement test was used here. The test consisted of 4 test items.

Table 4: Movement test sentences

<table>
<thead>
<tr>
<th>Category sequence</th>
<th>Word sequences</th>
<th>Grammatical in Grammar 1</th>
<th>Category sequence</th>
<th>Word Sequences</th>
<th>Grammatical in Grammar 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDEAB</td>
<td>JES SOT FAL KOF</td>
<td>DEABC</td>
<td>HOX</td>
<td>Hox JES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>REL ZOR TAF</td>
<td></td>
<td>DAZ NEB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TID LUM RUD</td>
<td></td>
<td>MER LEV</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TID ZOR RUD</td>
<td></td>
<td>MER NEB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two random orders were generated for each type (i.e., CDEAB and DEABC), resulting in four test samples (two G1-consistent and two G2-consistent samples). Each test sample was approximately 14.6 s in duration.
4.3. Procedure

We used the Head-turn Preference Procedure (Jusczyk & Aslin 1995; Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken 1995). Every infant accumulated a minimum of 70 s familiarization before proceeding on to the test phase. During the test phase, the four test samples were played in a random order. Half of the infants heard G1-consistent test sample first and the other half heard G2-consistent test sample first.

4.4. Results and discussion

Infants accumulated an average of 114 s acquisition time during the familiarization phase (range: 71–196 s). The mean looking time at either side during the test was 6.8 s. The standard deviation was 7.1 s. The data from the infants whose looking time during the test phase was over 2.5 standard deviations from the mean were not included in the analyses. The remaining infants in both conditions looked longer to the test samples that were inconsistent with their input grammar (mean = 8.50 s) than the test samples that were consistent with the input grammar (mean = 4.12 s). This difference was significant in a two-tailed Paired Samples t-test (t(1,12) = -2.42, p < 0.05). Eleven out of 13 infants had longer average looking times for the inconsistent samples.

![Figure 7: Results of the infant experiment](image-url)

Specifically, infants listened longer to test sentences in which a non-constituent in their input language was moved (“inconsistent” samples) than to test sentences in which constituents in their input language were moved...
consistent samples). This result suggests that infants showed a novelty preference and looked longer at the samples with novel structures. After less than 2 min of exposure, infants distinguished samples that were consistent with their input grammar from those that were inconsistent with their input grammar, as reflected by significantly longer looking times to the inconsistent samples. Finally, the results of this experiment indicate that 18-month-old infants can learn the phrase structure of an internally-structured artificial language on the basis of statistical distribution alone, consistent with and crucially extending the findings from Thompson & Newport (2007).

It should be noted that this infant experiment is a replication of the first adult experiment in which the input included movement sentences. So, we still do not know whether infants allow only constituents to undergo movement in the absence of movement sentences in the input. Thus, we are currently running experiments that exclude all movement sentences from the familiarization phase to see whether infants still infer the conclusion that only constituents can be moved.

5. SRN Simulations

Next, we present a series of neural network simulations with Simple Recurrent Networks (SRN) on the artificial language learning task. SRNs have been proposed to be able to learn a number of different aspects of human language, including syntax (Elman, et al. 1996; Elman 1991, 1993, Rohde & Plaut 1999). Recall that the results from the adult experiments showed that even in the absence of movement sentences in the input, adult participants still preferred sentences in which constituents were moved, in comparison to sentences in which non-constituents were moved. In other words, adults were able to generalize beyond the observed input. From these results, we inferred that that knowledge must have been known antecedently because it could not have arisen from the input. We carried out these simulations because SRNs have no inherent assumption about linguistic representation or structure, and we were interested in whether SRNs could learn to generalize beyond input.

5.1. Simulation 1

We used LENS (Rohde 1999), a simulation software, for all the simulations we report below. The input grammars for the network were identical to Grammars 1 and 2 that were used for the adult and infant experiments. In other words, the same 80 sentences that were used for adult experiments were chosen here as input. During training of the network, one word was presented at a time. The network had 35 hidden units. The activation of a previous word was copied into the context unit and when a word was presented as the input, the network’s task was to predict the following word.

One simulation run consisted of the training and the test, and 20 runs were carried out. For each run, either Grammar 1 or Grammar 2 was presented as
input during the training. At test, we presented the movement test sentences (the same 16 sentences used for the adult experiments) that were either consistent with Grammar 1 or Grammar 2. The prediction was that if the network did learn the artificial language, they should assign higher probabilities to the test sentences that are consistent with their input grammar.

We carried out a set of simulations with wide range of training parameters, and found a setting that resulted in the best performance (batch size: 39; learning rate: 0.001; weight updates: 500). This parameter setting was used for the following simulations.

The two input conditions (G1 and G2) are collapsed in the results reported below. The test items are labeled below as “consistent” and “inconsistent”. The consistent test items are sentences in which a constituent in the input grammar is moved, whereas the inconsistent test items are the sentences in which non-constituents in the input grammar is moved.

We first took the product of the probabilities that the network produced for each word in all the test sentences. Since it is the product of the probabilities, the numbers were extremely small, and we therefore computed the log of those numbers. Below, we report the mean of the log probability from 20 runs.

Simulation 1 included movement sentences in the input and at test, the network assigned higher probabilities to the consistent test items (mean log = -108.47) than to the inconsistent test items (mean log = -112.23). This difference was significant on the two-tailed t-test (p < 0.001).

This shows that the network correctly predicted the upcoming word when the test sentences were from their input grammar, while they performed worse when the test sentences were from the novel grammar.

Figure 8: Simulation 1

Figure 9: Simulation 2
5.2. Simulation 2

In Simulation 2, we excluded all movement sentences from the input, as we did in the second adult experiment. The test items were identical to Simulation 1. Again, we took the product of the probabilities that the network produced for each word in all the test sentences, then we converted them into log numbers.

The results showed that the network did not assign significantly higher probabilities to the consistent test items (mean log = -108.55) than to the inconsistent test items (mean log = -110.03; two-tailed t-test, \( p = 0.13 \)).

These results suggest that when the input includes movement sentences, the SRN can successfully learn the language and correctly assign higher probabilities to the sentences from their input grammar. However, when the input has no movement sentences, the network fails to distinguish the consistent test sentences and the inconsistent test sentences. This could be due to the fact that neither type of test items were seen in the input, and therefore, both were novel to the network. Moreover, if the network has no assumption about what kind of structure is linguistically valid or what kind of movement is allowed (e.g., movement of constituents), both test items are illicit according to what they learned. In other words, when faced with the unseen structure, the network seems to fail to extend their learning beyond the input. This is in contrast to our findings from the adult experiments, which showed that humans are able to generalize beyond the input.

6. Conclusion

We presented a series of experiments testing to see whether the nested phrase structure of an artificial language can be learned on the basis of transitional probabilities alone. We tested three different groups of subjects: human adults, infants and SRNs. The input grammars and the test items were the same for all subjects. All three groups performed well when the input included movement sentences. In those cases, when they were tested on movement sentences, both adults and the network favored the consistent test sentences, in which constituents as opposed to non-constituents were moved. Infants listened longer to the inconsistent test sentences showing a novelty preference.

The groups performed differently when the input did not include movement sentences. Adults still chose consistent test sentences over inconsistent test sentences even with the absence of movement in the input. However, the SRN could not discriminate two types of test items in this case. They performed at chance. What caused the difference in performance between adults and the network? We suggest that it might be the set of assumptions that the learner brings to the task. Human adults may have inherent knowledge about what kind of structures are representationally licensed (e.g., moved constituents), and when learning a new language, they may employ that knowledge. When faced with two novel sentences, even though neither was observed in the input, adults could correctly identify the consistent type, possibly because they had the constraint
that disallowed movement of non-constituents. On the other hand, SRNs do not have such inherent knowledge or constraints. They do not have any assumptions about what structure is licit or illicit. Therefore, when faced with two novel sentences, both are equally illicit according to what they have seen.

One of the main questions we were asking in the current paper was what is the form of knowledge that results from statistical learning. We presented two possibilities. One was that learners simply track the distributions and that they create the illusion of structure entirely based on the observed input. The other was that learners already come with a set of possible representations or structures and simply use the statistics to guide the learning. The results from the present paper suggest that the SRN behave as predicted by the first possibility. The product of the learning does not go beyond the observed input. On the other hand, humans performed as predicted by the second possibility. Humans generalized beyond the observed input, but within the limits of linguistic constraints. Thus, we suggest that not everything is learnable through the input and that the output of learning derives from an interaction between the input and an inherent representational system.

References


Maye, Jessica, Janet Werker, & LouAnn Gerken (2002). Infant sensitivity to
distributional information can affect phonetic discrimination. *Cognition, 82*, 101–
111.


Mintz, Toben (2003). Frequent frames as a cue for grammatical categories in child

Mintz, Toben, Elissa Newport, & Thomas Bever (2002). The distributional structure of

Morgan, James (1986). *From simple input to complex grammar*. Cambridge, MA: MIT
Press.

Morgan, James, Richard Meier, & Elissa Newport (1987). Structural packaging in the
input to language learning: Contributions of prosodic and morphological marking of

University of Maryland, College Park.

Press.

MA: Harvard University Press.

Redington, Martin, Nick Chater, & Steven Finch (1998). Distributional information: A

CMU-CS-99-164, Carnegie Mellon University, Department of Computer Science,
Pittsburgh, PA.

Rohde, Douglas & David Plaut (1999). Language acquisition in the absence of explicit

Saffran, Jenny, Richard Aslin, & Elissa Newport (1996). Statistical learning by eight-


Thompson, Susan & Elissa Newport (2007). Statistical learning of syntax: The role of
thorough probability. *Language Learning and Development, 3*, 1-42.

Yang, Charles (2006). *The infinite gift: How children learn and unlearn the languages of