Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge

Rebecca L. Gomez*, LouAnn Gerken

Speech and Hearing Sciences, University of Arizona, Tucson, AZ 85721, USA

Abstract

Four experiments used the head-turn preference procedure to assess whether infants could extract and remember information from auditory strings produced by a miniature artificial grammar. In all four experiments, infants generalized to new structure by discriminating new grammatical strings from ungrammatical ones after less than 2 min exposure to the grammar. Infants acquired specific information about the grammar as demonstrated by the ability to discriminate new grammatical strings from those with illegal endpoints (Experiment 1). Infants also discriminated new grammatical strings from those with string-internal pairwise violations (Experiments 2 and 3). Infants in Experiment 4 abstracted beyond specific word order as demonstrated by the ability to discriminate new strings produced by their training grammar from strings produced by another grammar despite a change in vocabulary between training and test. We discuss the implications of these findings for the study of language acquisition. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Grammar; Artificial learning; Infants

1. Introduction

Research shows that infants become sensitive to linguistic cues early in development. By 10 months of age, infants have acquired information specific to the inventory of target language segments (Werker and Tees, 1984), ordering of phones in

* Corresponding author. Tel.: +1-520-6263104; Fax: +1-520-6219901; e-mail: rgomez@u.arizona.edu

words (Jusczyk et al., 1993b, 1994), co-occurrence of acoustic cues at linguistic boundaries (e.g. Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989; Jusczyk et al., 1992; Gerken et al., 1994), language specific stress patterns of words (Jusczyk et al., 1993a; Newsome and Jusczyk, 1994), and phrasal patterns of grammatical morphemes (Shady et al., 1994, 1996; Shafer et al., 1998).

Despite a rapidly growing knowledge base on infant linguistic sensitivity, our understanding of the learning mechanisms involved in language acquisition has been limited. There are several reasons for this limitation. First, in the majority of studies, infants are presented with either natural passages versus passages that have been altered in some way, or with frequently occurring versus infrequently occurring linguistic patterns. However, various components of natural language are highly correlated, making it difficult to pinpoint the exact factors responsible for infants' discriminations. For example, determiners often occur at the beginnings of sentences or clauses and also precede nouns. Thus, inverting a determiner and noun not only disrupts word order, but also affects correlations of stress patterns at prosodic boundaries. Second, there is no way to control for prior learning in studies using natural language stimuli, thus making it impossible to watch and control learning 'in real time.' Third, until recently, there has been little attention to how infants might actually learn new structure and what types of structure they acquire. For instance, are the learning mechanisms involved in language acquisition specific to language or are they part of a general repertoire of learning tools (e.g. Rumelhart and McClelland, 1986; Pinker and Prince, 1988)? What role do general purpose frequency detectors or complex associative mechanisms (cf. Rumelhart and McClelland, 1986; Elman, 1990) play in language acquisition? How do such mechanisms factor into acquiring the hierarchical structure of language? A 'statistical learning mechanism' may be used by the learner to segment words in running speech (cf. Saffran et al., 1996), but does the learner also use such a mechanism to abstract form classes from word dependencies? In short, we know infants are sensitive to cues in language but we know little about how infants acquire or use this sensitivity in the service of language acquisition.

In order to circumvent such problems, our studies combine insights and methodology from two lines of research: one examining infant learning of artificial language materials (Morgan and Saffran, 1995; Mintz, 1996; Saffran et al., 1996; Echols et al., 1997) and another involving studies of artificial grammar learning by adults (Reber, 1967, 1969, 1989; Morgan and Newport, 1981; Morgan et al., 1987, 1989; Valian and Coulson, 1988; Valian and Levitt, 1996). The use of artificial language stimuli enables more precise control over the learning environment than is found using natural language. Such control enables systematic manipulation and testing of specific structural factors in language learning. Prior learning is also controlled, permitting the separation of effects due to artificial language exposure from those associated with prior knowledge.

Artificial stimuli have traditionally been used in adult studies, but recently, Saffran et al. (1996) used artificial stimuli to assess the mechanisms infants use to segment speech. Eight-month-olds were familiarized with 2 min of continuous

speech, consisting of four, trisyllabic nonsense words repeated in random order (e.g. *bidakupadotigolabubidaku...*). Infants were then tested on their ability to discriminate two of the familiarized words from two non-words. Familiar words and non-words were derived from the same syllable set, but differed in terms of transitional probabilities between syllable-pairs (with familiar words having mean transitional probabilities of 1.0 and non-words having mean transitional probabilities of zero).[1] Looking times resulted in a novelty preference for non-words over familiar words, suggesting that infants can make use of statistical properties of input to segment words in speech. In a second, more stringent examination of this finding, Saffran et al. demonstrated that infants could also discriminate familiar words from part-words (where the transitional probability between the first two syllables of the part-word was 0.33 and between the last two was 1.0). This result suggests that the statistical learning mechanism available to infants is powerful enough to discriminate the higher transitional probabilities occurring within words from the lower ones occurring between words in speech.

In another study using artificial stimuli. Mintz (1996) exposed 9-month-old infants to a learning environment consisting of three novel strings. Each string was composed of four unique words. Strings were presented in random order during acquisition for a total of 120 s (40 s for each string). At test, infants showed a listening preference for the familiarized strings versus strings consisting of the same words, but re-ordered so that they began and ended with words that had previously occurred in second and third positions.

Both Mintz (1996) and Saffran et al. (1996) provide important insights into infant language acquisition. However, in beginning to address the problem of syntax acquisition, a number of questions remain. First, both Mintz and Saffran tested infant memory for strings of elements presented during training. However, it is important to determine whether infants trained on a subset of grammatical strings will generalize to *new* grammatical strings they have never heard before. Second, given that word-order in English sentences is highly variable, will infants learn sequential dependencies on exposure to a grammar that allows words to occur in variable, rather than fixed, orders in sequence? A third and related point involves levels of transitional probabilities. Saffran et al. (1996) demonstrated that infants can use transitional probabilities to segment words when the transitional probability within words is 1.0, but we know little about the role transitional probabilities play at the level of syntax acquisition. Given that transitional probabilities between words in sentences is virtually never 1.0, can infants detect relations among units when transitional probabilities occur in a lower range? Finally, an important characteristic of syntax acquisition is the ability to abstract beyond specific utterances. We asked whether infants would show this ability.

The approach taken in the present experiments was to expose infants to auditory strings generated by a finite-state grammar. Such grammars have been used exten-

---

[1]The transitional probability of *y/x* is defined as the frequency of *xy* divided by the frequency of *x*. Transitional probability is a type of first-order dependency. We use the more general term in order to distinguish *n*-order dependencies.

sively in adult implicit learning research (Reber, 1967, 1989), therefore the learning that results is fairly well understood. Finite-state grammars are limited in terms of their generative capacities (Chomsky, 1957), but nevertheless are complex systems with a number of interesting properties. In fact, participants acquire structure at a variety of levels of complexity including, but not limited to, the first-order dependencies defined by the grammar (Gomez and Schvaneveldt, 1994; Gomez, 1997). These grammars also generate a variety of utterances. Thus, learners can be trained on a subset of grammatical strings and then tested for their ability to generalize to new strings.

During a typical adult artificial grammar learning study, participants are exposed to a subset of strings generated by a finite-state grammar (see Fig. 1), often in the context of a cover task. Participants are then given a surprise test to see whether they can discriminate new grammatical strings from ungrammatical ones. Participants show learning of the structure of the grammar even when learning is purely incidental (Reber, 1989). Learning also occurs rapidly, often with fewer than 10 minutes exposure to novel stimuli (e.g. see Altmann et al., 1995). Additionally, although participants can report partial strings (e.g. VOT JIC RUD; Gomez, 1997), they are typically unable to report the 'rules' of the grammar (Reber, 1989). Such performance shares similarities with language learning where children exhibit rule-like behavior without apparent knowledge of the rules of grammar. Finally, as noted above, learning occurs at various levels of complexity, ranging from knowledge of first-order dependencies, in which a given element in sequence is completely determined by its immediately preceding element (Perruchet and Pacteau, 1990; Gomez and Schvaneveldt, 1994) to second-order dependencies, involving two elements of prior context (Dienes et al., 1991; Gomez and Schvaneveldt, 1994). Studies also demonstrate transfer to stimuli with the same grammatical structure but instantiated in entirely new vocabulary (e.g. Reber, 1969; Mathews et al., 1989; Brooks and Vokey, 1991; Gomez and Schvaneveldt, 1994; Altmann et al., 1995; Knowlton and
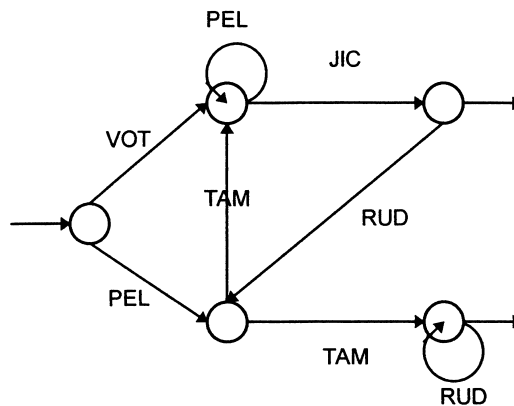

Fig. 1. Grammar 1. Finite-state grammar used to generate grammatical strings in Experiments 1–4. Strings are generated by starting at the leftmost state and then traversing links in the system in the direction of the arrows. An example of a grammatical string is VOT PEL JIC RUD TAM RUD.

Squire, 1996; Gomez, 1997). The ability to discriminate grammatical from ungrammatical instances in new vocabulary is taken as evidence that learners have abstracted some aspect of the grammar's structure (Reber, 1989).

We chose to conduct our research with 11- and 12-month-olds because of language changes taking place at approximately 10 months of age suggesting that infants are attaining a greater degree of linguistic maturity. As noted earlier, 10-month-olds' sensitivity to sound-contrasts has become specific to their target language (Werker and Tees, 1984). Infants of this age are also beginning to show sensitivity to grammatical morphemes thought to cue categories in language (Shady, 1996; Shafer et al., 1998).

Infants participated in two experimental phases. During acquisition, they were exposed to a subset of grammatical strings using the head-turn preference procedure (see Kemler Nelson et al., 1995). Infants were then presented with new grammatical and ungrammatical strings during test. The nature of the grammatical and ungrammatical strings was manipulated in four experiments. Given previous experiments showing longer orientation times toward natural compared to altered language stimuli, we expected infants to listen longer to grammatical strings than to ungrammatical ones (e.g. Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989; Jusczyk et al., 1992; Gerken et al., 1994; Mintz, 1996).

Experiment 1 investigated whether infants could extract endpoint information from a novel grammar by asking whether they would discriminate new grammatical strings from strings with illegal endpoints. Experiments 2 and 3 examined whether infants would extract first-order dependencies even when words occurred in variable positions in sequence. Experiment 2 examined whether infants were learning first-order word dependencies by testing their ability to discriminate new grammatical strings from ungrammatical ones containing pairwise violations of word order. Experiment 3 tested the generalizability of the learning observed in Experiment 2 by training infants on one of two grammars and testing them to see if they would discriminate new grammatical strings from their training grammar from strings produced by the other grammar. Experiment 4 assessed whether infants could abstract beyond specific word order by training them in one vocabulary and then exposing them to new vocabulary at test.

2. Experiment 1

2.1. Method

2.1.1. Participants

Infants were recruited via birth announcements published in the local newspaper. Sixteen American infants, approximately 12 months of age were tested. The infants had an average age of 361 days (range: 335–402). Thirteen additional infants were tested but not included for the following reasons: failed to look for an average of at least 3 s to each side during test ($n = 4$), cried ($n = 3$), low birthweight ($n = 3$; less than 5.5 lbs at birth), technical difficulties ($n = 3$). Because of changes in the first

Table 1
Acquisition strings used in Experiments 1 and 2

Set A	Set B
VOT PEL JIC	PEL TAM RUD RUD
PEL TAM PEL JIC	VOT JIC RUD TAM JIC
PEL TAM JIC RUD TAM RUD	VOT JIC RUD TAM RUD
PEL TAM JIC RUD TAM JIC	VOT PEL JIC RUD TAM
VOT PEL PEL JIC RUD TAM	PEL TAM PEL PEL PEL JIC

author's lab during the time in which the initial studies were carried out, seven infants were tested at New Mexico State University (henceforth Location 1) and nine were tested at the University of Arizona (Location 2).

2.1.2. Stimuli

Grammatical strings were generated by traversing the links in Grammar 1 (see Fig. 1) from left to right, starting at the leftmost position and ending at one of the rightmost positions. The loops in the grammar were optional. We set a limit of two repetitions on loops and generated a subset of grammatical strings ranging in length from three to six words. Grammar 1 (G1) generates 23 strings under these constraints.

2.1.2.1. Acquisition stimuli. Ten strings from G1 were used during acquisition. Although the 10 strings did not exhaust the grammatical set, they were chosen to represent all paths through the grammar. The strings were grouped into two sets of five (see Table 1). Three random orders were generated for each set, resulting in six acquisition samples. The mean transitional probability for acquisition strings is shown in Table 2.[2]

2.1.2.2. Test stimuli. Ten new grammatical and 10 ungrammatical strings were used during the test (see Table 3). The grammatical strings were also generated using G1 but were not in the acquisition set. The 10 ungrammatical strings were produced by interchanging the first and last words of each of the grammatical test strings. For instance, the grammatical string VOT PEL PEL PEL JIC was made ungrammatical by swapping VOT and JIC, resulting in JIC PEL PEL PEL VOT. Importantly, this procedure allowed us to match grammatical and ungrammatical strings in terms of word frequency and string length. Thus, any differences observed could not be attributed to these factors, but rather to infants' discriminations of legal and illegal endpoints.

---

[2]The equation shown in footnote 1 was used to compute mean transitional probabilities for training strings. For instance, the transitions VOT-PEL and PEL-JIC occurred with transitional probabilities of 0.6 and 0.38 in the training stimuli in Experiments 1 and 2. Averaging these values resulted in a mean transitional probability of 0.49 for the string VOT PEL JIC. Mean transitional probabilities for each of the 10 strings were averaged in order to obtain the mean transitional probability for the entire acquisition set.

Table 2
Mean transitional probabilities for acquisition and test strings in Experiments 1 and 2. Ranges are shown in parentheses

	Transitional Probabilities		
	Training	Grammatical test	Ungrammatical test
Experiment 1	0.42 (0.10–0.60)	0.42 (0.10–0.60)	0.24 (0–0.60)
Experiment 2	0.42 (0.10–0.60)	0.42 (0.10–0.60)	0.01 (0–0.60)

Note: Although infants were trained and tested on different grammatical strings, the transitional probabilities were identical for training and test stimuli.

The grammatical and ungrammatical strings were each grouped into two sets of five (see Table 3). Two random orders were generated for each grammatical and ungrammatical set, resulting in eight test samples (four grammatical and four ungrammatical). The mean transitional probabilities for grammatical and ungrammatical test strings is shown in Table 2.[3]

The 14 samples (six acquisition and eight test) were read by a female college professor who was blind to the conditions of the experiment. Strings were randomly ordered in the speaker's script. The speaker was instructed to read each string in an animated manner with a rising list intonation so that all words except the last received equal stress. The last word was spoken at a higher pitch and received more stress than the preceding words. The speaker was also asked to pause after each string for approximately 0.5 s before beginning the next string. The strings were recorded on audio tape. Each sample was approximately 18.5 s in duration. An example consisting of a random order of the five strings from Set A is PEL TAM PEL JIC | VOT PEL PEL JIC RUD TAM | VOT PEL JIC | PEL TAM JIC RUD TAM RUD | PEL TAM JIC RUD TAM JIC, where the '|' denotes a 0.5 s pause. Longer samples were then obtained by re-recording each of the original samples back-to-back (e.g. to obtain PEL TAM PEL JIC | VOT PEL PEL JIC RUD TAM | VOT PEL JIC | PEL TAM JIC RUD TAM RUD | PEL TAM JIC RUD TAM JIC | PEL TAM PEL JIC | VOT PEL PEL JIC RUD TAM | VOT PEL JIC | PEL TAM JIC RUD TAM RUD | PEL TAM JIC RUD TAM JIC). The resulting test trials were approximately 37 s in duration.

In order to guard against the possibility that infants might attend differentially to stimuli for some reason other than training, we compared grammatical and ungram-

---

[3]The equation shown in footnote 1 was used to compute mean transitional probabilities for test strings. These were computed relative to transitional probabilities in the training stimuli. For instance, the transitions PEL-TAM and TAM-RUD occurred with transitional probabilities of 0.38 and 0.27 in the training stimuli in Experiments 1 and 2. Averaging these values results in a mean transitional probability of .325 for the test string PEL TAM RUD. Mean transitional probabilities for each of the 10 grammatical strings were averaged in order to obtain the mean transitional probability for the entire set of grammatical test strings. The ungrammatical string TAM JIC RUD VOT yielded a mean transitional probability of $(0.36 + 0.54 + 0) \div 3 = 0.30$ (where the transitional probability for RUD-VOT was zero because this transition never occurred in the training strings). Mean transitional probabilities for each of the 10 ungrammatical strings were averaged in order to obtain the mean transitional probability for the entire set of ungrammatical test strings.

Table 3
Test strings used in Experiment 1

Grammatical	Ungrammatical
VOT JIC RUD TAM	TAM JIC RUD VOT
VOT PEL PEL JIC	JIC PEL PEL VOT
PEL TAM JIC RUD TAM	TAM TAM JIC RUD PEL
VOT PEL JIC RUD TAM JIC	JIC PEL JIC RUD TAM VOT
PEL TAM PEL JIC RUD TAM	TAM TAM PEL JIC RUD PEL
PEL TAM RUD	RUD TAM PEL
PEL TAM PEL PEL JIC	JIC TAM PEL PEL PEL
VOT PEL PEL PEL JIC	JIC PEL PEL PEL VOT
VOT JIC RUD TAM RUD RUD	RUD JIC RUD TAM RUD VOT
	RUD PEL JIC RUD TAM VOT

matical strings to determine whether they differed in average words per unit time and average frequency. Grammatical and ungrammatical test samples were similar in terms of average number of words per second (M = 1.36, SD = 0.01 vs. M = 1.35, SD = 0.03). The samples were also similar in terms of average frequency (M = 269 Hz, SD = 72 for grammatical strings vs. M = 264 Hz, SD = 64 for ungrammatical strings).

2.1.3. Apparatus

As noted above, data were collected in two locations. The apparatus used in Location 1 was manual whereas the Location 2 apparatus was automated. The mother and infant were in a soundproof booth in Location 2.

2.1.3.3. Location 1. The experiment was conducted in a three-sided plywood test booth, with panels of 4 × 6 ft on three sides and open at the back. There was a hole approximately 8 cm in diameter cut into the center board 10 cm above a green light which was mounted at the infant's eye level. A Sony compact video camera (model CCD-TR64) was aligned with the hole behind the center panel. Each of the side walls of the booth was mounted with a 7-inch Radio Shack Realistic loudspeaker and a red light at the infant's eye level. During the experiment, samples were played using a Radio Shack Realistic AM/FM cassette player (Modulair 850). There was a small 5 × 10 cm opening covered with one-way cellophane behind the center panel to permit an observer responsible for monitoring looking time to record the infant's head movements. This observer used a response box equipped with a timer and response button to record the duration of head turns. The observer recorded the duration of looking times on a coding sheet which indicated the target side for each trial. Another observer, who was responsible for controlling the lights and the onset of the stimulus samples, watched the infant on the video-recorder. The second observer used a control box which was equipped with a series of buttons for starting and stopping the flashing center and side lights. The second observer also manipulated a puppet between trials to help refocus the infant's attention at center and terminated a trial when the infant looked away for more than two seconds. Both

observers wore headphones and were uninformed as to which sample type was playing on any given trial. Sessions were videotaped to permit reliability checks on judgments about the duration of head turns.

2.1.3.4. Location 2. The audiotaped stimuli used in Location 1 were transferred to sound files using Signalize software running on a MAC OS computer. During the experiment, the computer controlled the presentation of the samples and recorded the observer's coding of the infant's responses. The audio output for the experiment was generated from the digitized waveforms of the samples and was fed to 7-inch Aiwa loudspeakers mounted on the side walls of a 9×9 ft soundproof testing booth. A Sony compact video camera (model CCD-TR64) was centered in front of the infant, above the infant's head level. An amber light was mounted directly in front of the infant at the infant's eye level. A red light was mounted on each of the side walls at the infant's eye level (directly under each of the speakers). In addition to the computer, a video monitor and response box were located outside the soundproof booth. The response box, which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of head turns, and terminated a trial when the infant looked away for more than 2 s. Information about the direction and duration of head turns and the total trial duration was stored in a data file on the computer. All sessions were videotaped to permit reliability coding. The observer viewed the infant on the video monitor rather than through one-way glass so that both observer and reliability coder used identical information when recording looking-times.

2.1.4. Procedure

We used the head-turn preference procedure (Kemler Nelson et al., 1995). Each infant was held on a caregiver's lap. The caregiver was seated in a chair in the center of the test booth. In order to maximize exposure to the acquisition stimuli, infants were exposed to two cycles of the six acquisition samples. Acquisition ended after 12 trials or when the infant appeared to lose interest in the task (as indicated by fussy behavior or failure to orient toward the sample for at least two seconds on three consecutive trials). Infants were then given a 5-min break. During the break the caregiver and experimenter played quietly with the infant. After the break, infants participated in an eight-trial test phase.

The acquisition phase familiarized the infant with both the samples and the procedure. During acquisition, the side of the initial trial was determined randomly, with subsequent trials occurring on alternate sides. During test, sides were determined randomly with the constraints that the initial trial was grammatical for half of the infants and ungrammatical for the other half and no more than three consecutive trials could occur on either side. Importantly, for each infant, both grammatical and ungrammatical strings were played from each side of the booth. Each infant participated in all eight test trials.

Each trial began by blinking the center light until the infant oriented in that direction. Then, the center light was extinguished and one of the side lights

began to flash. When the infant made a head turn of at least 30° in the direction of the flashing side light, the sample began to play and continued until its completion or until the infant failed to maintain the 30° head turn for 2 consecutive seconds (e.g. if the infant turned back to the center or the other side, looked at the mother, the floor or the ceiling). If the infant turned briefly away from the target by 30° in any direction, but for less than 2 s, and then looked back again, the time spent looking away was not included in the orientation time. The language sample terminated as soon as the time spent looking away from the target exceeded 2 s (even if this meant truncating a string in midstream). During the acquisition trials, the side light was extinguished when the sample began, but during the test trials the light remained on for the entire duration of the trial.[4] An observer recorded the duration of the infant's head turns using the response box and was uninformed as to which sample was played on a particular test trial. The infant's parent listened over Radio Shack Realistic Nova 14 headphones (Sony MDR 7502 headphones in Location 2) to a continuous presentation of acquisition samples. Parents reported that with this background they were unaware of the nature of the stimulus on any given trial.

We performed reliability coding of the videotapes made during each session to ensure that the observers did not bias the results. Videotapes of all 16 of the infants tested were available for reliability coding. A different observer from those who made the original observations viewed the videotapes with the soundtracks turned off and obtained a measure of looking time for each test trial. A comparison of live and videotaped observations (conducted according to the procedure outlined in Kemler Nelson et al., 1995) indicated high inter-observer agreement in looking times ($r_{14} = 0.97$). The discrepancy between times recorded by the live and videotape observers was less that 0.5 s on 65% of the trials and greater than 1 s on 15% of trials. Further examination of the latter type determined that there was no systematic tendency for the live observer to bias looking times in favor of the experimental hypothesis by, e.g. recording longer looking times for grammatical stimuli or shorter times for ungrammatical stimuli relative to the videotape observer ($ts \leq 0.15$, $Ps \geq 0.881$). The patterns of significant results for live and videotaped sessions were identical.

It should be noted that, since the infant's looking behavior determined the duration of exposure to the acquisition materials, no infant heard the entire set of acquisition samples. However, our randomization procedure maximized exposure to all of the acquisition strings. For instance, six of the 10 possible acquisition strings occurred first in the six acquisition samples. The remaining four strings occurred second. Infants could complete as many as 12 acquisition trials (two cycles of the six acquisition samples). However, each infant participated in at least six of the 12. Therefore, even infants with short listening times would have been exposed to all 10 acquisition strings. Given that the time to listen to 10 acquisition strings in Experiment 1 is approximately 37 s, an infant accumulating 74 s exposure during training

---

[4]Piloting from other laboratories (e.g. Kemler Nelson et al., 1995) convinced us that this was the best way to handle the lights during the procedure.

would have been exposed to each acquisition string approximately two times during training (i.e. 74 ÷ 37 = 2). An infant accumulating 41 s exposure (the minimum time recorded in Experiment 1), would have been exposed to each acquisition string approximately 1.08 times.

2.2. Results and discussion

The time that each infant oriented to the loudspeaker on each trial was recorded. Infants accumulated an average of 66 s acquisition time during the training phase (range: 41–114 s).[5] Looking times for each infant in each trial of the test phase were log transformed to limit individual differences in looking times.[6] Means were then computed separately across grammatical and ungrammatical trials for the first and second blocks, resulting in four means per infant. A repeated measures ANOVA with grammaticality status (grammatical vs. ungrammatical) and block (first four trials vs. second four) revealed a main effect of grammaticality, $F(1,15) = 9.09$, MSE = 1.28, $P = 0.009$. Average looking time to grammatical strings was 6.71 s (SD = 1.51) and to ungrammatical strings was 4.61 s (SD = 1.36). Ten of 16 of the infants had longer average looking times for the grammatical versus ungrammatical strings. There was also a main effect of block, $F(1,15) = 12.68$, MSE = 1.17, $P = 0.003$. Average looking time for the first four trials was 6.66 s (SD = 1.31) and 4.65 s (SD = 1.47) for the second four. A greater difference between looking times for grammatical and ungrammatical trials in the second compared to the first block would indicate that infants were learning during the test. There was no such interaction.

A Pearson product–moment correlation was performed to assess the relationship between total acquisition time and the difference in looking times to grammatical and ungrammatical trials. It revealed a positive, but non-significant value ($r_{14} = 0.20$).

The purpose of Experiment 1 was to determine whether one-year-olds can generalize to new strings and discriminate them from strings with illegal endpoints in a miniature artificial language. Infants distinguished new strings from those with endpoint violations, as reflected by significantly longer looking times to new grammatical compared to ungrammatical strings. Such learning occurred after very little exposure and was sustained over a 5-min delay. Correlations between total looking time during acquisition and the difference in looking times between grammatical and ungrammatical test strings were positive, but did not reach significance. There was no interaction of blocks with grammaticality, lending support to the argument

---

[5]Given that the time to listen to 10 acquisition strings in Experiment 1 is approximately 37 s, infants accumulating an average of 66 s exposure during training would have been exposed to each acquisition string approximately 1.78 times during training (i.e. 66 ÷ 37 = 1.78).

[6]We used the standard procedure, often used with adult reaction-time data, of applying log transformations to reduce variability due to outlying data points. A log transform was obtained for each observation before computing means on the log transformed data. The means reported in the text were converted to seconds from the log transformed means. Importantly, the patterns of results obtained from log transformed data were the same as those obtained using the raw observations.

that the familiarity preference for new grammatical strings can indeed be linked to training rather learning during the test. We will deal with this issue further in Experiments 3 and 4.

Given the ease with which infants acquired information about endpoints, our next objective was to determine whether 1-year-olds could also learn first-order dependencies in Grammar 1. The results reported by Saffran et al. (1996) certainly suggest that they should, but the transitional probabilities used in their studies were higher than those found in Experiment 1 or in natural language. Also, elements in their stimuli occurred uniquely and in fixed positions within words. Dependencies in natural language are much lower due to the fact that units occur more than once and in variable orders. Thus Experiment 2 investigated whether infants could learn grammatical word order when words occurred in variable positions in sequence and when first-order dependencies were lower than 1.0.

3. Experiment 2

3.1. Method

3.1.1. Participants

Infants were recruited in the same manner as in Experiment 1. Sixteen American infants of approximately 12 months of age were tested. The infants had an average age of 360 days (range: 305–407). Seven additional infants were tested but not included for the following reasons: failed to look for an average of at least 3 s to each side during test ($n = 2$), cried ($n = 2$), low birthweight or premature ($n = 1$), technical difficulties ($n = 2$).

3.1.2. Stimuli

3.1.2.1. Acquisition stimuli. The 10 acquisition strings used in Experiment 1 were used in Experiment 2. The strings were grouped in the same manner, with three random orders generated for each set of five strings, resulting in six acquisition samples. Each sample was used twice for a total of 12 acquisition trials. Because the acquisition strings in Experiments 1 and 2 were identical, the mean transitional probability for acquisition strings was the same as in Experiment 1. The mean transitional probability for acquisition strings is shown in Table 2.

3.1.2.2. Test stimuli. Ten grammatical and 10 ungrammatical strings were used during the test. The grammatical strings were the same as in Experiment 1. Ten ungrammatical strings were produced by beginning each string with a grammatical word, followed by a series of internal pairwise violations, and ending with a grammatical word (see Table 4). For instance, the string VOT TAM PEL RUD JIC begins and ends with grammatical words, but VOT TAM, PEL RUD, and RUD JIC are ungrammatical transitions. Grammatical and ungrammatical strings were matched as closely as possible in terms of word frequencies and string lengths

Table 4
Test strings used in Experiment 2. Both string types begin and end with legal words, but ungrammatical test strings contain illegal transitions (marked by asterisks)

Grammatical	Ungrammatical
VOT JIC RUD TAM	VOT*TAM PEL*RUD*JIC
VOT PEL PEL JIC	VOT*RUD*PEL JIC
PEL TAM JIC RUD TAM	PEL*RUD*JIC*PEL TAM
VOT PEL JIC RUD TAM JIC	VOT*TAM PEL JIC*PEL*RUD
PEL TAM PEL JIC RUD TAM	PEL*RUD*JIC*JIC*TAM*TAM
PEL TAM RUD	PEL*VOT*TAM
PEL TAM PEL PEL JIC	PEL*RUD*JIC*PEL TAM
VOT PEL PEL PEL JIC	VOT*RUD TAM*TAM
VOT JIC RUD TAM RUD RUD	PEL*VOT*RUD*PEL JIC*JIC
VOT PEL JIC RUD TAM RUD	VOT*RUD*PEL TAM PEL JIC

so that any differences observed could not be attributed to these factors, but instead would be due to infants' discriminations of legal and illegal transitions. The mean transitional probabilities for grammatical and ungrammatical test strings is shown in Table 2.

As in the first experiment, grammatical and ungrammatical strings were grouped into sets of five. Two random orders were generated for each grammatical and ungrammatical string set, resulting in eight test samples (four grammatical and four ungrammatical). The samples were recorded in the same manner as in Experiment 1.

Acoustic analyses showed that grammatical and ungrammatical test samples were equated in terms of average number of words per second (M = 1.43, SD = 0.07 vs. M = 1.43, SD = 0.05) and average frequency (M = 256 Hz, SD = 61 for grammatical strings vs. M = 254 Hz, SD = 59 for ungrammatical strings).

3.1.3. Apparatus

Six of the infants were tested using the apparatus in Location 1; the other 10 were tested at Location 2.

3.1.4. Procedure

The procedure during acquisition and test was identical to Experiment 1. Infants participated in all eight test trials. We performed the same reliability procedures to ensure that the observer who timed the duration of trials did not bias the results. Videotapes of 15 of the 16 participants tested were available for reliability coding in which a different observer from the one who made the original observations viewed the videotapes with the soundtracks turned off. Inter-observer agreement in looking times was high ($r_{13} = 0.99$). The discrepancy between times recorded by the live and videotape observers was less that 0.5 s on 61% of the trials and greater than 1 s on 17% of trials. Further examination of the latter type determined that there was no systematic tendency for the live observer to bias looking times in favor of the

experimental hypothesis by recording longer looking times for grammatical stimuli or shorter times for ungrammatical stimuli relative to the videotape observer ($ts \leq 0.29$, $Ps \geq 0.778$). The patterns of significant results for live and videotaped sessions were identical.

3.2. Results and discussion

The time that each infant oriented to the loudspeaker on each trial was recorded. Infants accumulated an average of 76 s acquisition time during the training phase (range: 40–117 s).[7] Looking times for each infant in each trial of the test phase were log transformed. Means were computed separately across grammatical and ungrammatical trials in a manner identical to the first experiment. A repeated measures ANOVA with grammaticality (grammatical vs. ungrammatical) and block (first four trials vs. second four) was performed on the log transformed data. The analysis revealed a main effect of grammaticality, $F(1,15) = 5.56$, $MSE = 1.27$, $P = 0.032$, but no block effect or interaction. Average looking time to grammatical strings was 8.00 s (SD = 1.57) and to ungrammatical strings was 5.99 s (SD = 1.34). Twelve of the 16 infants had longer average looking times for the grammatical versus ungrammatical strings.

A Pearson product-moment correlation was performed to assess the relationship between total acquisition time and the difference in looking times to grammatical versus ungrammatical strings. It revealed a positive, but non-significant value ($r_{14}) = 0.10$.

The high agreement in words per second and average frequency (reported in the methods section) indicates that infants' preference for grammatical over ungrammatical strings was not due to prosodic differences between string types. Nevertheless, as a further check, we performed an experiment with eight normal hearing adults using low-pass filtered versions of the stimuli (this eliminates phonetic and phonotactic information while preserving the prosodic features of the stimuli). If prosodic differences between grammatical and ungrammatical strings were contributing to performance, participants should be able to discriminate grammatical from ungrammatical strings under these conditions.

Participants listened to an audio-tape of the acquisition samples that had been used with the infants (resulting in three minutes exposure to the acquisition stimuli). They were then told that the sample they had listened to consisted of strings generated using a set of complex rules. Participants were instructed to determine whether each of eight new samples was 'similar' or 'not similar' to the strings to which they had just listened. Participants were warned that the strings were all low-pass filtered and so would sound 'somewhat different,' but that four of the samples followed the same rules as the acquisition strings and four contained violations of the rules. Participants listened to each of the eight low-pass filtered samples and

---

[7]Given that the time to listen to 10 acquisition strings in Experiment 2 is approximately 35 s, infants accumulating an average of 76 s exposure during training would have been exposed to each acquisition string approximately 2.17 times during training.

indicated their judgment by marking 'S' for similar or 'NS' for not similar on a numbered score sheet. They were asked to repeat the test phase with stimuli that had not been low-pass filtered as a reliability check on the general procedure.

Mean proportion correct was 0.52 (SD = 0.14) on the low-pass filtered stimuli and did not differ significantly from chance (0.50), $t(7) = 0.38$, $P = 0.715$. In contrast, participants scored well above chance on discrimination of grammatical and ungrammatical samples that were not low-pass filtered. The mean proportion correct was 0.83 (SD = 0.11), $t(7) = 20.54$, $P < 0.001$, demonstrating that adults had acquired enough information to identify new grammatical strings. This result, in combination with the fact that discrimination was at chance with the low-pass filtered stimuli, suggests that first-order dependencies between words, and not prosodic cues were contributing to grammaticality performance.

Experiment 2 extended the findings from Experiment 1 by asking whether infants were processing first-order dependencies in addition to endpoint information. Infants in Experiment 2 discriminated new strings from those with internal pairwise violations as reflected in significantly longer looking times to new grammatical compared to ungrammatical strings. Furthermore, infants appear to be processing first-order dependencies over a range of values. Such learning is remarkable given that it occurs with little exposure and is retained over a 5-min delay. Additionally, learning does not appear to be limited to the exact strings presented during training, as evidenced by the fact the infants show the discrimination on new strings at test.

In order to eliminate the possibility that the grammaticality effects we observed in our original studies were due, not to learning, but to infants recognizing that new grammatical strings were, in some way, more systematic than ungrammatical strings, and to show that learning would generalize beyond our specific grammar, we next trained infants on one of two grammars (the grammar shown in Fig. 1 and the one shown in Fig. 2). Both grammars began and ended with the same words, but
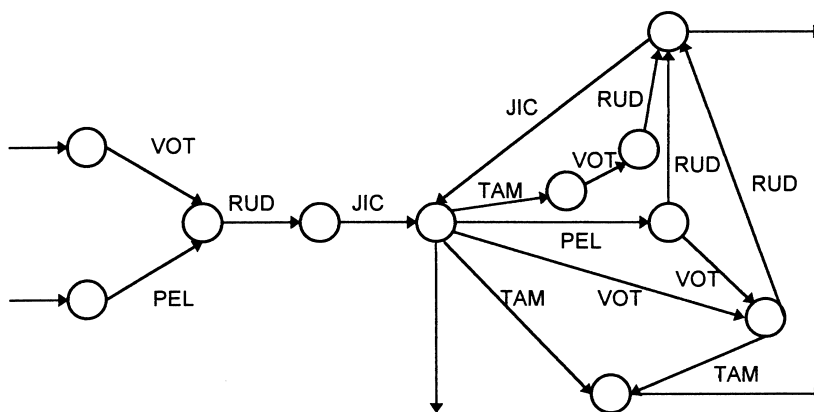

Fig. 2. Grammar 2. Finite-state grammar used to generate grammatical strings in Experiment 3.

differed in terms of internal word ordering. Infants were then tested to see if they would discriminate new strings from their training grammar from strings from the other grammar regardless of which grammar they were exposed to during training.

4. Experiment 3

4.1. Method

4.1.1. Participants

Infants were recruited in the same manner as in the first two experiments. Sixteen American infants of approximately 11.5 months of age were tested. The infants had an average age of 343 days (range: 307–385). Nine additional infants were tested but not included for the following reasons: cried ($n = 3$), low birthweight or premature ($n = 3$), belonged to a family with a history of language impairment ($n = 1$), technical difficulties ($n = 2$).

4.1.2. Stimuli

Twenty strings were generated from Grammar 1 (G1) shown in Fig. 1 and 20 from Grammar 2 (G2) shown in Fig. 2. The strings from each grammar were grouped into four sets of five (Sets A–D, shown in Table 5). The order of strings in each set was then randomized six times, resulting in 48 stimulus samples. The strings were recorded by one of the authors (L.A.G.). The six randomly ordered strings in one set were recorded before going on to another set. Sets were blocked in this manner to allow the speaker to obtain sufficient practice with each set. The last two samples read for each set were used as test stimuli. The six strings in each set were edited using Signalize software to ensure that each string was separated by 0.5 s. Each sample was recorded back-to-back in the same manner as in the first two experiments in order to produce sufficiently long listening trials.

Sets were counter-balanced (see Table 6) so that four infants were trained on six samples each from Sets A and B (from G1), and four were trained on six samples each from Sets C and D (from G1). Similarly, four infants were trained on Sets A and B (from G2) and four were trained on Sets C and D (from G2). The two remaining sets from the training grammar (10 strings total) and two sets from the other grammar were then used as test stimuli. For instance, infants trained on G1, Sets A and B, were tested on Sets C and D from G1 and G2 (see Table 6). G1 and G2 samples were equated in terms of average words per unit time (M = 1.48, SD = 0.02 for G1 vs. M = 1.47, SD = 0.03 for G2) and average frequency (M = 272 Hz, SD = 63 for G1 vs. M = 277 Hz, SD = 76 for G2). The mean transitional probabilities for training and test strings are listed in Table 7.

As in the previous experiments, there were 12 possible acquisition samples and 8 test samples. The grammaticality status of the sample presented on the first test trial was counterbalanced across infants so that half of the infants received a sample from their training grammar and half received a sample from the other grammar.

4.1.3. Apparatus

All 16 infants were tested using the apparatus at Location 2. Thus, all children were tested in a soundproof booth and the procedure was entirely automated.

4.1.4. Procedure

The procedure during acquisition and test was identical to that used in Experiments 1 and 2. Infants participated in all eight test trials. We performed the same reliability procedures to ensure that the observer who timed the duration of trials did not bias the results. Videotapes of 15 of the 16 participants tested were available for reliability coding. Inter-observer agreement in looking times was high ($r_{13} = 0.92$). The discrepancy between times recorded by the live and videotape observers was less that 0.5 s on 61% of the trials and greater than 1 s on 18% of trials. Further examination determined that there was no systematic tendency for the live observer

Table 5

Training and test stimuli used in Experiment 3. Both grammars begin and end with the same words, but none of the internal transitions are shared

Grammar 1

Set A	Set B
VOT PEL JIC	VOT JIC RUD TAM
VOT PEL PEL PEL JIC	PEL TAM RUD RUD
PEL TAM JIC RUD TAM JIC	PEL TAM PEL PEL JIC
VOT PEL JIC RUD TAM RUD	VOT PEL JIC RUD TAM
PEL TAM PEL JIC RUD TAM	PEL TAM JIC RUD TAM RUD

Set C	Set D
PEL TAM RUD	VOT PEL PEL JIC
PEL TAM JIC RUD TAM	PEL TAM PEL JIC
VOT PEL PEL JIC RUD TAM	PEL TAM RUD RUD RUD
VOT PEL JIC RUD TAM JIC	VOT JIC RUD TAM JIC
VOT JIC RUD TAM RUD RUD	PEL TAM PEL PEL PEL JIC

Grammar 2

Set A	Set B
PEL RUD JIC	VOT RUD JIC TAM
VOT RUD JIC VOT TAM	PEL RUD JIC PEL RUD
PEL RUD JIC VOT RUD JIC	PEL RUD JIC VOT RUD
PEL RUD JIC PEL VOT TAM	VOT RUD JIC VOT RUD JIC
VOT RUD JIC TAM VOT RUD	VOT RUD JIC PEL VOT RUD

Set C	Set D
VOT RUD JIC	PEL RUD JIC TAM
VOT RUD JIC VOT RUD	PEL RUD JIC VOT TAM
VOT RUD JIC PEL VOT TAM	VOT RUD JIC PEL RUD
PEL RUD JIC PEL VOT RUD	PEL RUD JIC PEL RUD JIC
VOT RUD JIC PEL RUD JIC	PEL RUD JIC TAM VOT RUD

Table 6
Counterbalancing design used in Experiments 3 and 4

Group	Training	Grammatical Test	Ungrammatical Test
1	G1 – Sets AB	G1 – Sets CD	G2 – Sets CD
2	G1 – Sets CD	G1 – Sets AB	G2 – Sets AB
3	G2 – Sets AB	G2 – Sets CD	G1 – Sets CD
4	G2 – Sets CD	G2 – Sets AB	G1 – Sets AB

to bias looking times in favor of the experimental hypothesis by, e.g., recording longer looking times for grammatical stimuli or shorter times for ungrammatical stimuli relative to the videotape observer ($ts \leq 0.06$, $Ps \geq 0.95$). The patterns of significant results for live and videotaped sessions were identical.

4.2. Results and discussion

The time that each infant oriented to the loudspeaker on each trial was recorded. Infants accumulated an average of 77 s acquisition time during the training phase (range: 50–127 s).[8] Looking times for each infant on each trial of the test phase were log transformed. Means were computed separately across grammatical and ungrammatical trials in a manner identical to that used in Experiments 1 and 2. A mixed design ANOVA with grammar (G1 vs. G2) and training set (Sets AB vs. CD) as between-subjects variables, and grammaticality (grammatical vs. ungrammatical) and block (first four trials vs. second four) as within-subject variables, was performed on the transformed data. The analysis revealed a main effect of grammaticality, $F(1,12) = 13.51$, MSE = 1.13, $P = 0.003$, but no effect of grammar, set, block, or any interactions involving these variables. Infants listened longer to new samples from their training grammar (M = 8.82, SD = 1.45) compared to samples from the other grammar (M = 6.45, SD = 1.35), regardless of which grammar they were trained on. Thirteen of 16 of the infants showed this effect. A Pearson product-moment correlation was performed to assess the relationship between total acquisition time and the difference in looking times to grammatical and ungrammatical stimuli. It failed to reach significance ($r_{14} = 0.03$).

Infants were trained on one of two grammars in Experiment 3. Training and test sets were counterbalanced in order to ensure that the learning effects observed in the first two experiments were not due to extraneous factors, such as arbitrary preferences for certain patterns or sound sequences specific to Grammar 1. The results of Experiment 3 showed that performance differences observed in the first two experiments were indeed due to learning. Grammars 1 and 2 produced strings of similar lengths and began and ended with the same words, but differed in terms of internal word ordering. Nevertheless, infants showed a listening preference for strings pro-

---

[8]Given that the time to listen to 10 acquisition strings in Experiment 3 is approximately 34 s, infants accumulating an average of 77 s exposure during training would have been exposed to each acquisition string approximately 2.26 times during training.

Table 7

Mean transitional probabilities for acquisition and test strings in Experiment 3. Ranges are shown in parentheses

Training grammar	Transitional probabilities		
	Training	Grammatical test	Ungrammatical test
G1 – Sets AB	0.44 (0.10 – 0.80)	0.40 (0.10 – 0.80)	0
G1 – Sets CD	0.38 (0.20 – 0.60)	0.39 (0.20 – 0.60)	0
G2 – Sets AB	0.61 (0.17 – 0.83)	0.60 (0.17 – 0.83)	0
G2 – Sets CD	0.61 (0.17 – 0.80)	0.60 (0.17 – 0.80)	0

Note: Transitional probabilities for test items were computed relative to transitional probabilities found in the training set (see footnote 3). Legal word pairs in G1 and G2 were non-overlapping, therefore transitional probabilities for ungrammatical tests strings were all zero.

duced by their training grammar compared to strings produced by the other grammar, regardless of which grammar they were trained on. As in the first two experiments, such learning occurred after brief exposure and was retained despite a 5-min delay. Test strings were similar, but not identical to training strings. Thus, infants showed learning of first-order dependencies despite the fact that the context in which these dependencies occurred varied between training and test.

An important characteristic of syntax acquisition is the ability to abstract beyond specific word pairs (or first-order dependencies). Thus, a remaining question is whether infants will show this ability. A convenient feature of finite-state grammars is that they can be instantiated in any vocabulary. This means that learners can be trained on strings such as FIM-SOG-FIM-FIM-TUP and then tested on PEL-TAM-PEL-PEL-JIC where every occurrence of FIM is mapped to PEL, every occurrence of SOG to TAM, and every occurrence of TUP to JIC. Grammatical structure is held constant across training and test instances despite the change in vocabulary. Because test strings are instantiated in new vocabulary, learners can not distinguish the two grammars based on transitional probabilities between remembered word pairs. Thus, discrimination in new vocabulary would suggest that learners have abstracted some aspect of grammatical structure above and beyond pairs of specific elements.

Adults readily transfer to new vocabulary (Reber, 1969; Mathews et al., 1989; Gomez and Schvaneveldt, 1994; Altmann et al., 1995; Knowlton and Squire, 1996; Gomez, 1997). However, transfer cannot be explained in terms of knowledge of specific word pairs. Gomez and Schvaneveldt (1994) trained adults on entire strings or on pairs of elements comprised from grammatical strings. Participants trained on pairs of elements acquired knowledge of first-order dependencies, but were not able to use this information to detect violations of second-order dependencies or to transfer to new vocabulary. In contrast, participants trained on strings acquired knowledge of first- and second-order dependencies and transferred this knowledge to new vocabulary. Like the adults in Gomez and Schvaneveldt's study, infants should not transfer to new vocabulary if their learning is limited to knowledge of specific word pairs, but they should transfer if they have abstracted beyond specific word order. We trained infants in one vocabulary and transferred them to new vocabulary at test as a means of evaluating this conjecture.

5. Experiment 4

5.1. Method

5.1.1. Participants

Infants were recruited in the same manner as in the first two experiments. Sixteen American infants of approximately 11.5 months of age were tested. The infants had an average age of 342 days (range: 331–357). Eleven additional infants were tested but not included for the following reasons: cried ($n = 4$), low birthweight or premature ($n = 1$), failed to look for an average of at least three seconds to each side during the test ($n = 1$). Additionally, extensive pilot testing led us to set a minimum of 50 s total listening time during acquisition. Five infants failed to meet this preset criterion.

5.1.2. Stimuli

The strings were identical to those used in Experiment 3 with the exception that acquisition strings were presented in the vocabulary JED, FIM, TUP, DAK, SOG and test strings were presented in the vocabulary VOT, PEL, JIC, RUD, TAM. The acquisition strings for Experiment 4 were constructed from the strings shown in Table 5 by mapping every occurrence of VOT to JED, PEL to FIM, JIC to TUP, RUD to DAK and TAM to SOG. For instance, application of this mapping to the string VOT RUD JIC TAM resulted in the training string JED DAK TUP SOG.

As in Experiment 3, strings from each grammar were grouped into four sets of five. The order of strings in each set was then randomized six times, resulting in 48 stimulus samples. Sets were counter-balanced in the same manner as Experiment 3. Because the new vocabulary was used during training, the test strings in Experiment 4 were identical to those used in Experiment 3. An important difference, however, was that rather than talker-produced strings, strings in Experiment 4 were constructed from spoken word tokens using Sound Edit software (word tokens were recorded by L.A.G.). Word tokens were used as an added control to ensure that there were no talker-induced differences in individual strings or in strings particular to one grammar or the other (e.g. the same occurrence of each word token was used for both G1 and G2 strings). G1 and G2 strings were equated in terms of average words per unit time (M = 1.39, SD = 0.01 for G1 vs. M = 1.42, SD = 0.01 for G2) and average frequency (M = 294 Hz, SD = 57 for G1 vs. M = 298 Hz, SD = 57 for G2). Mean transitional probabilities for the training stimuli were the same as those found in Experiment 3 (see Table 7). Because of the change in vocabulary between training and test, transitional probabilities for grammatical and ungrammatical test stimuli were zero.

In the first three experiments, listening times to the first test trial were long compared to subsequent trials. In order to reduce the variability contributed by the first trial, infants in Experiment 4 first received a warm-up trial, followed by eight test trials (the number used in Experiments 1–3). As in the previous experiments the grammaticality status of the first test trial was counterbalanced across infants. In Experiment 4, half of the infants received a warm-up trial from the

acquisition phase (but in new vocabulary) and half received a warm-up trial from the other grammar (in new vocabulary). This trial was not included in the analyses.[9] After the warm-up trial, infants participated in eight test trials (none of which were identical to the warm-up trial). The first test trial was also counterbalanced so that half of the infants who received a warm-up trial from their training grammar also received the first test trial from their training grammar and half received the first test trial from the other grammar. The same procedure was followed for infants receiving their warm-up trial from the other grammar.

5.1.3. Apparatus

All 16 infants were tested using the apparatus at Location 2. Thus, all children were tested in a soundproof booth and the procedure was entirely automated.

5.1.4. Procedure

The procedure during acquisition and test was identical to that used in Experiments 1–3. Infants participated in all nine trials at test (one warm-up trial and eight test trials). We performed the same reliability procedures to ensure that the observer who timed the duration of trials did not bias the results. Videotapes of 15 of the 16 participants tested were available for reliability coding. Inter-observer agreement in looking times was high ($r_{13} = 0.99$). The discrepancy between times recorded by the live and videotape observers was less than 0.5 s on 81% of the trials and greater than 1 s on 7% of trials. Further examination determined that there was no systematic tendency for the live observer to bias looking times in favor of the experimental hypothesis by, e.g. recording longer looking times for grammatical stimuli or shorter times for ungrammatical stimuli relative to the videotape observer (ts ≤ 0.54, Ps ≥ 0.61). The patterns of significant results for live and videotaped sessions were identical.

5.2. Results and discussion

The time that each infant oriented to the loudspeaker on each trial was recorded. Infants accumulated an average of 72 s acquisition time during the training phase (range: 51–108 s).[10] Looking times for each infant on each trial of the test phase were log transformed. Means were computed separately across grammatical and ungrammatical trials in a manner identical to that used in the previous experiments. A mixed design ANOVA with grammar (G1 vs. G2) and training set (Sets AB vs. CD) as between-subjects variables, and grammaticality (grammatical vs. ungrammatical) and block (first four trials vs. second four) as within-subject variables, was

---

[9]Analyses of Experiments 1–3 with the first trial removed resulted in the same pattern of effects. However, removing the first trial led to a significant reduction in variability. Because of this, we adopted the procedure of including a warm-up trial in subsequent experiments.

[10]Given that the time to listen to 10 acquisition strings in Experiment 4 is approximately 35.5 s, infants accumulating an average of 72 s exposure during training would have been exposed to each acquisition string approximately 2.03 times during training.

performed on the transformed data. It revealed a main effect of grammaticality, $F(1,12) = 5.16$, MSE = 0.20, $P = 0.042$, but no effect of grammar, set, block, or any interactions involving these variables. Infants listened longer to new samples from their training grammar (M = 6.67, SD = 1.42) compared to samples from the other grammar (M = 5.18, SD = 1.32), regardless of which grammar they were trained on. Twelve of 16 of the infants showed this effect. A Pearson product-moment correlation was performed to assess the relationship between total acquisition time and the difference in looking times to grammatical and ungrammatical stimuli. It failed to reach significance ($r_{14} = -0.13$).

The purpose of Experiment 4 was to determine if infant learning is limited to specific pairwise associations. Previous research shows that knowledge of isolated word pairs is insufficient to produce transfer across vocabularies (Gomez and Schvaneveldt, 1994). Therefore, we trained infants in one vocabulary and transferred them to new vocabulary at test. Infants were able to discriminate new strings from their training grammar from strings from the other grammar, despite the change in vocabulary. This finding demonstrates that they were extracting information in the form of larger units, perhaps involving second-order dependencies or series of first-order dependencies, but not limited to isolated word pairs. This finding also suggests that one-year-olds are abstracting some aspects of the structure of their training grammar. Adult research on transfer has led to several proposals regarding the basis of abstraction. These will be addressed briefly in Section 6.

6. General discussion

The present experiments tested infants' ability to learn a miniature artificial grammar. Infants in all four experiments were exposed to a subset of word strings generated by a finite-state grammar. Infants were able to discriminate new grammatical from ungrammatical strings after less than two minutes exposure to the grammar. Learning was all the more remarkable given that it was retained over a 5-min delay.

Previous studies (Mintz, 1996; Saffran et al., 1996) provide important insights into infant language acquisition, but we were particularly interested in extending such research to questions relevant to syntax acquisition. First, in all four experiments, infants were not tested on their memory for old grammatical strings, but instead were asked to generalize to new instances of learned structure. That is, the test strings in the present experiments were similar, but not identical to strings presented during training. Word pairs occurred regularly in grammatical strings, but in slightly different positions in sequence depending on the string. Infants were able to distinguish grammatical from ungrammatical structure in the context of such changes suggesting flexibility in their learning. Second, infants in our experiments were exposed to grammars characterized by variable word order. That is, words occurred in multiple and variable positions in sequence depending on the series of state transitions taken through the grammar. Previously reported studies have investigated sequential learning of uniquely occurring units. One impli-

cation of using uniquely occurring units is that word order is fixed. Another implication is that first-order dependencies are always 1.0. However, words rarely occur in fixed order in sentences of natural language and first-order dependencies are rarely 1.0. Thus, we also investigated whether infants could learn first-order dependencies in ranges more likely to be found in natural language. For instance, the mean transitional probability for grammatical test items in Experiments 1 and 2 was 0.42, and in Experiment 3 was 0.395 for G1 and 0.60 for G2. Despite these added complexities, infants demonstrated rapid learning of first-order dependencies in their training grammar. Finally, the ability to abstract beyond specific word pairs is a critical feature of syntax acquisition (Chomsky, 1957). Previous research (Gomez and Schvaneveldt, 1994) has shown that knowledge of isolated word pairs is insufficient for producing transfer to new vocabulary. Therefore, infants who were learning isolated word pairs should not perform transfer. Because test strings were instantiated in new vocabulary, transitional probabilities between word pairs was zero. Infants transferred easily, demonstrating that they were abstracting beyond the transitional probabilities holding between particular words in the grammar.

Specifically, Experiment 1 asked whether infants could generalize to new instances of learned structure by discriminating *new* grammatical strings from those with illegal endpoints. The answer was affirmative. Experiment 2 asked whether infants would distinguish n*ew* grammatical strings from strings with legal endpoints, but with internal pairwise violations of grammatical word order. Again, the answer was affirmative. Experiment 3 trained infants on one of two grammars and tested their ability to discriminate *new* strings from their training grammar from strings produced by the other grammar. Both grammars began and ended with the same words, but word ordering differed within strings. Infants showed the discrimination regardless of their training grammar (demonstrating that the previous results were not due to idiosyncrasies of the grammar used in the first two experiments). Experiment 4 trained infants in one vocabulary and transferred them to new vocabulary at test. Infants discriminated *new* strings from their training grammar from strings from the other grammar, *despite the change in vocabulary*, demonstrating that they were not only learning specific word pairs, but were also extracting information of a more abstract form.

Adult research identifies two possible explanations for this abstraction. Given that finite-state grammars allow paths to cycle, one possibility is that learners abstract frames with repeating elements, i.e. of the form _x(x)_ _ _ or _x_ x_ _ (Mathews and Roussel, 1997). Patterns fitting the frames are then used to identify legal strings. A similar proposal is that learners perform abstract analogies between specific training and test items by noting similarities in patterns of repeating elements (Brooks and Vokey, 1991). Although abstraction of such patterns does not reflect abstraction of the grammar per se, it does reflect generalization of a complex form (namely, abstraction of identity structure).

According to a very different view (Altmann et al., 1995; Dienes et al., in press) learners use a complex association mechanism (Elman, 1990) to abstract the sequential structure of the grammar. By this view, learners acquire series (or trajectories) of

associations from information about only two elements of sequential input at a time. They then induce mappings between vocabularies during transfer (Altmann et al., 1995; Dienes et al., in press). One prediction of this model is that repeating elements should have no special status. Thus transfer should occur just as readily after exposure to grammars without repetitions as it does after exposure to grammars with them.

We are currently performing a series of experiments designed to determine which of these accounts of transfer best explains abstraction in adult and infant artificial grammar learning (Gomez and Gerken, 1998; Gomez et al., in press). Regardless of the mechanism used during transfer, both explanations are consistent with the view that infants are sensitive to a range of information in strings, including endpoints, internal word pairs, and information of a more abstract form.

In closing we would like to address two remaining issues. First, infants appear to acquire first-order dependencies readily and with ease (Experiments 1–3; Newsome and Jusczyk, 1994; Saffran et al., 1996), but what does this tell us about language acquisition? A statistical learning mechanism capable of processing first-order dependencies almost certainly plays a role in word segmentation. First-order dependencies between syllables can be used to identify words. Saffran et al. (1996) have clearly demonstrated this. Such learning might also factor into acquiring longer-distance dependencies, such as those between specific auxiliaries (e.g. 'was') and their requisite morphemes (e.g. 'ing') (e.g. Santlemann and Jusczyk, 1997). However, the next step of inducing grammatical categories must certainly involve some degree of abstraction beyond specific word pairs. For instance, given enough experience children may learn to associate 'was' and 'ing' with particular words (that happen to be verbs), but inducing the grammatical category of a novel word like 'dax' in the sentence 'Mary was daxing' almost certainly requires some form of abstraction. A statistical learning mechanism that processes transitional probabilities among linguistic cues may also play a role in segmenting linguistic units larger than words (e.g. clauses and phrases), but again, syntactic use of these units must certainly involve abstraction beyond specific utterances. The degree to which first-order dependencies factor into abstraction processes has yet to be determined. Experiment 4 suggests that infants are abstracting some aspect of sequential structure, but future artificial grammar learning studies should be designed to test abstraction processes that might be involved in acquiring knowledge of grammatical categories. We are in the process of testing such learning with infants and adults.

Second, the present experiments attest to the feasibility of training infants using artificial grammars, but given that such grammars differ vastly from natural language, how can we ensure that the learning observed is representative of human language acquisition? It may be impossible to circumvent this problem entirely, but we can certainly take precautions in our methodology. In using this approach, it is important to take care in designing experiments that capture key linguistic phenomena. If we can isolate a phenomenon of interest experimentally, then we can determine the effect of a wide range of manipulations on the learning observed. We can compare findings obtained using artificial grammars to those found using natural language. One prediction is that the same developmental trends should emerge when

testing similar phenomena using natural and artificial languages. For instance, the finding of Werker and Tees (1984) that infants lose the ability to recognize non-native contrasts by 10 months of age suggests that infants by this age are beginning to form more abstract representations of phonological categories. Findings reported by Shady (1996) and Shafer et al. (1998) suggest that infants begin abstracting syntactic structure soon afterward. Both Shafer et al. and Shady report that infants notice violations of grammatical morphemes in novel sentences by 11 months of age, but Shafer et al. report that infants do not show this ability a month earlier. If these developmental benchmarks are indeed reflecting the development of rudimentary abstraction abilities, then we might predict that although 11-month-olds would acquire syntax-like categories in an artificial grammar, 9-month-olds would not. Additional information might be obtained by comparing the developmental trends observed in language to those observed for the development of general cognitive abilities (e.g. see Lalonde, 1989; Lalonde and Werker, 1995).

In sum, infants exposed to a subset of the strings produced by an artificial language generalize to novel grammatical strings with a minimum of exposure. Infants also show remarkable abstraction abilities, as manifested in the ability to generalize their newfound knowledge to novel strings instantiated in new vocabulary. Exposing infants to artificial grammars such as these promises to extend our knowledge about infant language learning abilities in important ways.

Acknowledgements

References

Altmann, G.T.M., Dienes, Z., Goode, A., 1995. Modality independence of implicitly learned grammatical knowledge. Journal of Experimental Psychology: Learning. Memory and Cognition 21, 899–912.

Brooks, L.R., Vokey, J.R., 1991. Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. (1989). Journal of Experimental Psychology: General 120, 316–323.

Chomsky, N., 1957. Syntactic structures. Mouton, The Hague.

Dienes, Z., Broadbent, D., Berry, D., 1991. Implicit and explicit knowledge bases in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory and Cognition 17 (5), 875–887.

Dienes, Z., Altmann, G.T.M., Gao, S.-J. Mapping across domains without feedback: a computational model. Cognitive Science, in press.

Echols, C.H., Crowhurst, M.J., Childers, J.B., 1997. The perception of rhythmic units in speech by infants and adults. Journal of Memory and Language 36, 202–225.

Elman, J.L., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Gerken, L.A., Jusczyk, P.W., Mandel, D.R., 1994. When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. Cognition 51, 237–265.

Gomez, R.L., 1997. Transfer and complexity in artificial grammar learning. Cognitive Psychology 33, 154–207.

Gomez, R.L., Gerken, L.A., 1998. Determining the basis of abstraction in infant artificial grammar learning. Poster presented at the 13th Biennial International Conference on Infant Studies.

Gomez, R.L., Gerken, L.A., Schvaneveldt, R.W. The basis of transfer in artificial grammar learning. Memory and Cognition, in press.

Gomez, R.L., Schvaneveldt, R.W., 1994. What is learned from artificial grammars? Transfer tests of simple association. Journal of Experimental Psychology: Learnin, Memory and Cognition 20, 396–410.

Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Wright Cassidy, K., Druss, B., Kennedy, L., 1987. Clauses are perceptual units for young infants. Cognition 26, 269–286.

Jusczyk, P.W., Cutler, A., Redanz, N.J., 1993a. Infants' preference for the predominant stress pattern of English words. Child Development, 64, 675–687.

Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., Jusczyk, A.M., 1993b. Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language, 32, 402–420.

Jusczyk, P.W., Hirsh-Pasek, K., Kemler Nelson, D.G., Kennedy, L., Woodward, A., Piwoz, J., 1992. Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology 24, 252–293.

Jusczyk, P.W., Luce, P.A., Charles-Luce, J., 1994. Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language 33, 630–645.

Kemler Nelson, D., Hirsh-Pasek, K., Jusczyk, P.W., Wright Cassidy, K., 1989. How prosodic cues in mothers might assist language learning. Journal of Child Language 16, 53–68.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., Gerken, L.A., 1995. The head-turn preference procedure for testing auditory perception. Infant Behavior and Development 18, 111–116.

Knowlton, B.J., Squire, L.R., 1996. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. Journal of Experimental Psychology: Learning, Memory and Cognition 22, 169–181.

Lalonde, C.E., 1989. An investigation of the relations among object search skills, cross-language speech perception, and visual categorization in infancy. Unpublished Master's Thesis, University of British Columbia.

Lalonde, C.E., Werker, J.F., 1995. Cognitive influences on cross-language speech perception in infancy. Infant Behavior and Development 18 (4), 459–475.

Mathews, R.C., Buss, R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R., Druhan, B., 1989. Role of implicit and explicit processes in learning from examples: a synergistic effect. Journal of Experimental Psychology: Learning, Memory and Cognition 15, 1083–1100.

Mathews, R.C., Roussel, L.G., 1997. Abstractness of implicit knowledge: a cognitive evolutionary perspective. In: Berry D.C. (Ed.), How Implicit is Implicit Learning? Oxford University Press, Oxford, pp. 13–47.

Mintz, T.L., 1996. The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories. Unpublished doctoral thesis. University of Rochester, Rochester, NY.

Morgan, J.L., Meier, R.P., Newport, E.L., 1987. Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. Cognitive Psychology 19, 498–550.

Morgan, J.L., Meier, R.P., Newport, E.L., 1989. Facilitating the acquisition of syntax with transformational cues to phrase structure. Journal of Memory and Language 28, 360–374.

Morgan, J.L., Newport, E.L., 1981. The role of constituent structure in the induction of an artificial language. Journal of Verbal Learning and Verbal Behavior 20, 67–85.

Morgan, J.L., Saffran, J.R., 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development 66, 911–936.

Newsome, M.R., Jusczyk, P.W., 1994. Do infants use stress as a cue in segmenting fluent speech? In: MacLaughlin, D., McEwen, S. (Eds.), Proceedings of the 19th Boston University Conference on Language Development, Cascadilla Press, Boston, MA.

Perruchet, P., Pacteau, C., 1990. Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General 119, 264–275.

Pinker, S., Prince, A., 1988. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. Cognition 28, 73–93.

Reber, A.S., 1967. Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior 77, 317–327.

Reber, A.S., 1969. Transfer of syntactic structure in synthetic languages. Journal of Experimental Psychology 81, 115–119.

Reber, A.S., 1989. Implicit learning and tacit knowledge. Journal of Experimental Psychology: General 118, 219–235.

Rumelhart, D.E., McClelland, J.L., 1986. On learning the past tenses of English verbs. In: Rumelhart, D.E. McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 1: Foundations, MIT Press, Cambridge, MA.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by eight-month-old infants. Science 274, 1926–1928.

Santlemann, L., Jusczyk, P.W., 1997. What discontinuous dependencies reveal about the size of the learner's processing window. Proceedings of the 21st Boston University Conference on Language Development, Cascadilla Press, Boston, MA, pp. 506–514.

Shady, M., Gerken, L.A., Jusczyk, P.W., 1994. Some evidence of sensitivity to prosody and word order in 10-month-olds. In: MacLaughlin, D. McEwen, S. (Eds.), Proceedings of the 19th Boston University Conference on Language Development, Cascadilla Press, Boston, MA.

Shady, M.E., 1996. Infants' sensitivity to function morphemes. Unpublished doctoral dissertation, State University of New York at Buffalo, NY.

Shafer, V.L., Shucard, J.L., Gerken, L.A., Shucard, D.W., 1998. 'The' and the brain: an electrophysiological examinations of infants' responses to familiar and unfamiliar function morphemes. Journal of Speech, Language and Hearing Research 41, 874–886.

Valian, V., Coulson, S., 1988. Anchor points in language learning: the role of marker frequency. Journal of Memory and Language.

Valian, V., Levitt, A., 1996. Prosody and adults' learning of syntactic structure. Journal of Memory and Language 35, 497–516.

Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7, 49–63.