CHAPTER 36

Statistical learning in infant language development

Rebecca Gómez

36.1 Introduction

Children learn language over such a short span of time with such seeming ease that many have assumed they must master language by means of a language-specific device (Chomsky, 1965; Wexler and Culicover, 1980). Such a device is thought to narrow down the search space of possible languages by limiting the parameters children will detect in language (Hyams, 1986; Manzini and Wexler, 1987; Roeper and Williams, 1987). Earlier in the twentieth century the idea that children might learn by means of associative principles was widespread (Skinner, 1957), but this was discounted because learning mechanisms were thought to be inadequate for acquiring the complex structure of human language (Chomsky, 1959). At that time, prosodic and distributional variations in language were not recognized as useful cues to linguistic structure, and thus it was assumed that language input could provide little useful information for learning (cf. Bates and Elman, 1996). However, learning approaches are gaining in credibility, spurred by the successes of neural network models (Elman et al., 1996), but also by discoveries with infant humans, suggesting a greater role for learning in language development than was previously assumed, and raising questions about the nature and extent of learning. The present chapter documents the infant language learning literature to date.

An approach to learning, taken for many years with adults but only in the past decade with infants, is to familiarize them with artificial languages. Although natural language stimuli are indispensable for investigating many questions in language development, they have drawbacks for assessing learning. For one, natural language is too familiar. Exposure to natural language is extensive and ongoing, making it impossible to know when and how infants learn about some particular linguistic structure. With novel stimuli, scientists can establish the point in developmental time when infants are able to detect a particular structure (e.g. Thiessen and Saffran, 2003). They can determine how much exposure is necessary for learning and what form exposure should take. And, they can ascertain how one learning experience affects another (e.g. Lany et al., 2005). Second, natural language is rich in correlated cues making it difficult to isolate the cause of learning. Is one particular cue sufficient for learning or are multiple cues needed (Billman, 1989)? If so, are some cues more powerful than others (Thiessen et al., 2005)? Artificial languages provide a useful tool for controlling prior learning and for manipulating specific variables of interest. This approach has resulted in a wealth of findings regarding the learning capabilities of children. Such capabilities enable infants to tune into the predominant auditory patterns in their native language and reflect learning across a range of linguistic tasks, including the acquisition of phoneme categories (Maye et al., 2002; Maye et al., 2005) and phonotactic dependencies (Chambers et al., 2003), phonological inference

(Gerken, 2004), segmentation of words (Saffran et al., 1996; Saffran, 2001), and acquisition of finite-state grammars (Gómez and Gerken, 1999; Saffran and Wilson, 2003; Saffran, Hauser, et al., 2005). Infant learners are also able abstract higher-order relations necessary for acquiring syntax-like structure (Gerken et al., 2005; Gómez and Lakusta, 2004).

Artificial language learning typically involves a two- to three-minute familiarization phase followed by a test. Most studies employ a two-language design so that half of the infants are exposed to Language A and half to Language B. At test, infants are exposed to strings from both languages so that what is grammatical for one group is ungrammatical for the other (Language A strings violate the constraints of Language B and vice versa). This ensures that the structure of the languages, instead of something idiosyncratic about the sound tokens used in one language or the other, is responsible for learning. Infants are tested for the amount of time they attend to different stimulus types. If learning has occurred, a group of infants will listen differentially to strings that conform to their training language vs. strings that do not (see Kemler Nelson et al., 1995). Adult learners can be used to obtain detailed information that cannot easily be gotten from infants. And comparisons between infants and adults are useful for exploring developmental differences in learning. Adults provide judgements about legal and illegal strings or are tested on their ability to produce legal sentences.

36.2 Statistical learning

Infant artificial language learning has become synonymous with statistical learning because of the emphasis in much of the work on learning statistical regularities. However, not all cases of artificial language learning entail learning statistical structure. For instance, some learning requires generalization of relational patterns (Gerken, 2004; Gerken et al., 2005; Gómez and Gerken, 1999; Gómez and Lakusta, 2004; Lany et al. 2005; Marcus et al., 1999).

To clarify, statistical structure abounds in the flow of information in the sensory world. Whether auditory, visual, or tactile, stimulation is rich in frequent associations and patterns over basic units. Statistical learning is thought to involve computations based on basic units. In the case of auditory stimulation, for instance, the basic units of computation are various linguistic or acoustic sounds. Statistical structure can take many forms, including the frequency of individual units of sound, frequency of co-occurrence, or the transitional probability of one unit given another. Co-occurrence frequency is defined as the frequency of two units occurring together. Transitional probability is the probability of the occurrence of one unit given another (the probability of Event B given Event A is the joint probability of Event A and Event B divided by the probability of Event A). There are other forms of statistical structure, but common to all forms is the requirement that units occur with some regularity that lends itself to mathematical description (and also presumably computation).

In a seminal paper, Saffran et al. (1996; see also Aslin et al., 1998) showed that eight-month-old humans can track transitional probabilities in sequences of syllables to discover word boundaries. They tested this by exposing infants to continuous streams of four randomly ordered three-syllable words (e.g. *tupiro*, *golatu*) such that individual syllables occurred with identical frequency. The only cues to word boundaries were the lower transitional probabilities occurring for syllables spanning words compared to the higher probabilities of syllables within words. Take a phrase like *naughty puppy*. The syllable transition in *naugh-ty* has a higher transitional probability than the transition *ty-pu* because *naugh* in the word *naughty*, is more likely to predict *ty* than *ty* is to predict *pu*. Infants in the Saffran et al. studies were able to use the differences in transitional probabilities within- vs. between-words to identify word boundaries in running speech.

This research was important for showing that very young infants can track complex sequential structure. It also raised more general questions about the role of learning in language acquisition, such as whether learning was limited to word segmentation or whether it applied to other kinds of language structure. A flurry of studies followed; and although much work remains to be done, learning appears to extend into in a range of linguistic domains.

Learning is key in the formation of speech categories (Maye et al., 2002; 2005), the ability to track phonotactic sequential structure (Chambers et al., 2003), the abstraction of phonological rules (Gerken, 2004), and the identification of word-like units in speech (Saffran, 2001). Infants are also able to acquire rudimentary syntactic structure in the form of adjacent and remote sequential dependencies (Gómez and Gerken, 1999; Gómez, 2002; Gómez and Maye, 2005; Saffran and Wilson,

2003), abstractio
(Gerken, 2006;
Marcus et al., 19
phosyntactic categ
2005; Gómez and
also shows how l
bilistic input (G
Hudson et al., 200
prior experience t
cult patterns from
2005; Lany and G
and other findings

36.3 Phonol

36.3.1 Discrimi
sounds

By 8–10 months o
narrow the invent
speech in their na
Tees, 1984). This fin
experience-depend
change? One theory
on growing perce
(e.g. Mackain, 198
However, a puzzle i
infants are unable t
pairs in an experime
months (Swingley a
2002), even though
sounds themselves
Werker, 1997). A po
use statistical info
phoneme categories
tokens in phonetic
along a dimension,
dom. Rather, it pat
tokens of one catego
along a dimension
category. In contra
organize according t
Maye et al. asked h
different characterist
ability to distinguish

They familiarized
one of two distributi
on a /da/–/ta/ conti
rated /d/ in day and t
in stay). Infants hav
discrimination (Peg
perception of speech
exposure to a unim
interfere with discrin
to a bimodal distribut
pattern of findings w
that sensitivity to the

, including the frequency
of sound, frequency of
transitional probability of
nother. Co-occurrence
the frequency of two units
nsitional probability is the
rrence of one unit given
ty of Event B given Event
lity of Event A and Event
ability of Event A). There
istical structure, but com-
e requirement that units
larity that lends itself to
ion (and also presumably

affran et al. (1996; see also
ved that eight-month-old
nsitional probabilities in
to discover word bound-
s by exposing infants to
f four randomly ordered
e.g. *tupiro*, *golatu*) such
s occurred with identical
ues to word boundaries
onal probabilities occur-
ning words compared to
ties of syllables within
like *naughty puppy*. The
ugh-ty has a higher tran-
an the transition *ty-pu*
word *naughty*, is more
an *ty* is to predict *pu*.
et al. studies were able to
transitional probabilities
words to identify word
speech.
portant for showing that
track complex sequential
l more general questions
ing in language acquisi-
learning was limited to
r whether it applied to
ge structure. A flurry of
although much work
rning appears to extend
istic domains.
he formation of speech
, 2002; 2005), the ability
c sequential structure
03), the abstraction of
Gerken, 2004), and the
d-like units in speech
s are also able to acquire
structure in the form of
sequential dependencies
, 1999; Gómez, 2002;
)5; Saffran and Wilson,

2003), abstraction of sequential patterns (Gerken, 2006; Gómez and Gerken, 1999; Marcus et al., 1999), and learning of morphosyntactic category relations (Gerken et al., 2005; Gómez and Lakusta, 2004). The research also shows how learners capitalize on probabilistic input (Gómez and Lakusta, 2004; Hudson et al., 2005a; 2005b) and how they use prior experience to bootstrap learning of difficult patterns from simpler ones (Lany et al., 2005; Lany and Gómez, forthcoming). These and other findings are detailed below.

36.3 Phonological learning

36.3.1 Discrimination of speech sounds

By 8–10 months of age infants have begun to narrow the inventory of sounds that count as speech in their native language (Werker and Tees, 1984). This finding is widely recognized as experience-dependent. What is the process of change? One theory is that change is dependent on growing perceptual sensitivity to words (e.g. Mackain, 1982; Werker and Pegg, 1992). However, a puzzle is presented by the fact that infants are unable to distinguish minimal word pairs in an experimental setting before seventeen-months (Swingley and Aslin, 2000; Werker et al., 2002), even though they distinguish the speech sounds themselves much earlier (Stager and Werker, 1997). A possible solution is that infants use statistical information to home in on phoneme categories (Maye et al., 2002). Although tokens in phonetic space may vary acoustically along a dimension, such variation is not random. Rather, it patterns bimodally such that tokens of one category are closer to one another along a dimension than tokens from another category. In contrast, within-category tokens organize according to a unimodal distribution. Maye et al. asked how distributions with these different characteristics might influence infants' ability to distinguish speech contrasts.

They familiarized six-month-old infants with one of two distributions of eight speech sounds on a /da/–/ta/ continuum (the voiced unaspirated /d/ in day and the voiceless unaspirated /t/ in stay). Infants have been shown to make this discrimination (Pegg and Werker, 1997); but if perception of speech sounds is malleable then exposure to a unimodal distribution should interfere with discrimination, whereas exposure to a bimodal distribution should preserve it. This pattern of findings would support the proposal that sensitivity to the frequency distributions of speech sounds is instrumental in learning. Infants in both unimodal and bimodal conditions heard the same eight tokens along the continuum. Those in the bimodal distribution condition heard tokens near the end (2 and 7) most frequently, whereas infants in the unimodal distribution condition heard middle tokens 4 and 5 most often. After familiarization, infants were tested on their ability to discriminate alternating tokens (the endpoints 1 and 8) from non-alternating ones (repeats of tokens 3 or 6). Only infants in the bimodal condition discriminated alternating from non-alternating tokens. Familiarization with the bimodal distribution apparently led infants to retain two categories of speech sounds, whereas exposure to the unimodal distribution resulted in the formation of one. In a subsequent study with eight-month olds, Maye et al. (2005) found that exposure to a bimodal frequency distribution could enable detection of an initially undetectable contrast. Thus, it appears that the frequency characteristics of speech sounds can both blur distinctions between previously known categories and enable the formation of new ones. Is statistical learning instrumental in the more basic task of learning phonotactic structure?

36.3.2 Learning phonotactic regularities

Phonotactic regularities are the allowable positions of speech segments relative to each other in words. For instance, English words can end with an /ng/ sequence (such as in the word *sing*), but such a sequence cannot begin words. Such regularities differ from language to language and thus must be learned. This is a non-trivial task requiring infants to remember the positions of individual speech segments across innumerable syllables and words (Chambers, 2004). We know infants distinguish frequent from infrequent phonotactic patterns in their native language by nine months of age (Friederici and Wessels, 1993; Jusczyk et al., 1993; Jusczyk et al., 1994), but until recently we have not known how rapidly new phonotactic regularities are acquired or whether learning will generalize.

Chambers et al. (2003) found that sixteen-month-old infants rapidly learn consonant position regularities involving syllable onsets and codas. Infants had to learn that CVC[1] words began with one set of five consonants, but

---

[1] C = consonant, V = vowel.

ended with another set of five. Infants were exposed to twenty-five words in a familiarization phase less than two minutes long. At test, they discriminated novel words with the familiar pattern from ones with an unfamiliar one (constraints were reversed such that illegal items began with final consonants and ended with initial ones). In another experiment, Chambers (2004) found that the same-age infants could extend newly learned phonotactic regularities to syllables with a different vowel, suggesting that they are able to form rule-like generalizations such as "/b/ is an onset." Chambers (2004) has also found that rapid learning extends beyond the first-order onsets of syllables: sixteen month olds were able to learn second-order dependencies in which the position of a consonant, whether first or last in a syllable, depended on an adjacent vowel.

Such experiments provide insights into the kinds of phonotactic regularity infants can learn, and may shed light on the formation of phonological units. Yang (2004) has pointed out that to apply statistical operations, learners must know what kinds of unit are relevant (whether syllables or some other kind of structure). These studies raise the intriguing possibility that some aspects of syllable structure could be acquired with the same principles that are instrumental in phonotactic learning. Although the infant learners were beyond the age at which children begin showing knowledge of syllable structure, Chambers (2004) has begun testing learning of phonotactic structure in younger infants.

36.3.3 Phonological generalization

An interesting question arises as to whether learning is limited to previously encountered syllables and segments or whether learners are sensitive to higher-level abstract patterns. For instance, languages vary in the range of syllable structures they allow. English permits a range from V, CV, and CVC to complex combinations such as CCCVCCC, whereas other languages allow less variation. Presumably such variations are learned. How malleable is such learning? Saffran and Thiessen (2003) investigated whether they could change infant preferences for phonological patterns in their linguistic input by exposing nine month olds to words conforming only to CVCV or CVCCVC patterns (both of which occur regularly in English). Infants discriminated novel words exhibiting the familiarization pattern from words exhibiting the other pattern after a brief exposure, showing that their preferences for common syllable patterns are easily altered.

In another study, Saffran and Thiessen (2003) exposed the same-age infants to words with CVC syllables with onsets and codas of +Voicing/−Voicing (*dakdot*) vs.−Voicing/+V voicing (*todkad*). As in the previous study, infants discriminated words with the familiar pattern from words with the other pattern, even though both patterns can occur in English. A third study suggested that infants had abstracted the pattern at the level of features (e.g. +Voicing/−Voicing) rather than at the level of having learned only the positions of specific segments. This is an important indication that learning extends to higher-level phonological relations.

Maye et al. (forthcoming) have also asked whether infants are encoding information at the level of abstract features (e.g. voiced vs. voiceless). To address this, Maye et al. familiarized eight month olds with a bimodal phonetic distribution emphasizing a contrast at one place of articulation (/g/ versus /k/) and tested them on discrimination of a contrast at another place of articulation (/d/ versus /t/). Both the /g/–/k/ and the /d/–/t/ distributions varied along a continuum of voicing. Infants were able to make the discrimination, showing that by 8 months they are encoding speech sounds according to the abstract feature of voicing. This suggests that infants are not merely learning individual phonetic contrasts but are generalizing patterns at a higher level.

In a yet more demanding demonstration of phonological generalization, Gerken (2004) investigated nine-month-old infants' ability to learn words generated by principles of metrical stress. Principles of stress assignment, applied to words of three to five syllables in length, were ordered along a hierarchy of constraints. For instance, Constraint A states that two stressed syllables can not occur in sequence, whereas Constraint B states that heavy syllables (such as those with a consonant ending) should be stressed. Because A outranks B, the word *TON ton do RE mi* does not stress the second syllable, even though Constraint B calls for stress. Infants were familiarized with words reflecting a system of four constraints (A–D). Importantly, they were never exposed to a word requiring the transitive inference that A outranks D, but they were familiarized with words for which B outranked C and C outranked D. At test they generalized to novel words requiring the transitive inference that A outranks D. One interpretation of the findings is that infants are abstracting a system of symbolic propositions and making a transitive inference; however, Shultz and

---

Gerken (2005)
encoder network
eralization, dem
achieved by a fu
Importantly, the
obtained from n
that infants can l
tionships found
This is an impor
questions regard
language studies
The literature
rapid learning at
relations (Cham
2003) and for ph
et al., 2002). Lear
across a range o
2004; Gerken, 20
Saffran and Thie
peting stress cues
ies have looked at l
as frequency or s
learning is taken
such as when cue

36.4 Word

36.4.1 Compet cues

A problem faced b
of words in runni
cult by the fact th
demarcated by pa
Segmentation is
including transiti
(Saffran et al., 199
(Mattys and Ju
(Johnson and Ju
1999), co-articula
2001), and highly f
2005; Jusczyk, 19
begun to investiga
How does statistic
phonological cues?
nent cue such as

---

[2] The learning algorith
cade-correlation. Cas
recruit hidden units
simulating various de
pared to back-propagat
weights in static networ

[3] See Curtin and Werker
ical development for a
on infant word segment

ffran and Thiessen (2003)
e infants to words with
onsets and codas of
dakdot) vs.–Voicing/+V
in the previous study,
words with the familiar
ith the other pattern, even
s can occur in English.
ested that infants had
at the level of features
g) rather than at the level
the positions of specific
mportant indication that
igher-level phonological

oming) have also asked
coding information at the
s (e.g. voiced vs. voiceless).
aye et al. familiarized
a bimodal phonetic dis-
a contrast at one place of
/k/) and tested them on
ntrast at another place of
/t/). Both the /g/–/k/ and
ns varied along a contin-
ts were able to make the
ng that by 8 months they
sounds according to the
icing. This suggests that
learning individual pho-
generalizing patterns at a

nding demonstration of
ization, Gerken (2004)
th-old infants' ability to
by principles of metrical
ss assignment, applied to
syllables in length, were
rchy of constraints. For
states that two stressed
r in sequence, whereas
t heavy syllables (such as
ant ending) should be
tranks B, the word *TON*
stress the second syllable,
t B calls for stress. Infants
words reflecting a system
–D). Importantly, they
o a word requiring the
t A outranks D, but they
words for which B out-
ked D. At test they gener-
requiring the transitive
ks D. One interpretation
infants are abstracting
ropositions and making
; however, Shultz and

Gerken (2005) showed that a feedforward encoder network[2] was capable of the same generalization, demonstrating that learning can be achieved by a fundamentally associative system. Importantly, the stress assignment rules were obtained from natural language, demonstrating that infants can learn a complex system of relationships found in languages of the world. This is an important demonstration because of questions regarding the relevance of artificial language studies to natural language.

The literature reviewed to this point shows rapid learning at the level of segments and their relations (Chambers, 2004; Chambers et al., 2003) and for phoneme discrimination (Maye et al., 2002). Learning is also complex, occurring across a range of generalizations (Chambers, 2004; Gerken, 2004; Maye et al., forthcoming; Saffran and Thiessen, 2003) and despite competing stress cues (Gerken, 2004). All these studies have looked at learning of individual cues such as frequency or stress; but what happens when learning is taken to another level of difficulty such as when cues of different types compete?

36.4 Word segmentation[3]

36.4.1 Competition of distributional cues

A problem faced by all children is identification of words in running speech. This is made difficult by the fact that words are not consistently demarcated by pauses (Liberman et al., 1967). Segmentation is aided by a variety of cues, including transitional probabilities of syllables (Saffran et al., 1996), word-boundary phonetics (Mattys and Juszczyk, 2001), word stress (Johnson and Jusczyk, 2001; Jusczyk et al., 1999), co-articulation (Johnson and Jusczyk, 2001), and highly familiar words (Bortfeld et al., 2005; Jusczyk, 1997), but studies have only begun to investigate the interplay of such cues. How does statistical information interact with phonological cues? Does sensitivity to a prominent cue such as stress precede sensitivity to transitional probabilities, or is the reverse the case?

Infants are sensitive to stress cues early on. By 7.5 months, infants more easily detect words in running speech that adhere to the predominant trochaic strong–weak stress pattern of English than words with an iambic weak–strong pattern (Jusczyk et al., 1999). Indeed, when presented with passages containing an iambic word such as *guitar* followed by an unstressed syllable such as *is* infants prefer to listen to *taris* over *guitar*, suggesting they are using a strong–weak segmentation strategy to identify words (Cutler and Norris, 1988; Jusczyk et al., 1999).

To use such a strategy, infants need to identify the predominant stress pattern of their native language (whether trochaic or iambic), but this is itself dependent on knowing some words (Thiessen and Saffran, 2003). One way out of this problem is if infants can use statistical information to segment words before they become sensitive to stress (Thiessen and Saffran, 2003). Infants might only segment a subset of words in their native language, but if they then form a generalization for the predominant stress pattern, they will have an additional cue for finding words in running speech. Interestingly, when stress and transitional probabilities are pitted against each other, eight month olds rely on stress (Johnson and Juszyk, 2001). What was not known, until recently, is whether infants would favor transitional probabilities over stress at an earlier age.

Thiessen and Saffran investigated this question by exposing six and nine month olds to a continuous stream of disyllabic words with trochaic or iambic stress. Infants of both ages were then tested to see if they would discriminate words from part-words. Infants are known to segment words based on trochaic, but not iambic, stress by 7.5 months of age, and do not appear to rely on iambic stress for word segmentation until 10.5 months (Juszcyk et al., 1999). Words consisted of syllables that were uniquely paired and thus had high transitional probabilities. Part-words consisted of syllables that co-occurred as frequently as the syllables in words, but contained syllables crossing word boundaries. Thus, the transitional probability of one syllable following another was lower in part-words than in words. Thiessen and Saffran reasoned that if infants favor statistical information over stress they should discriminate words and part-words the same, regardless of whether they were in the trochaic or iambic condition; but if they are already sensitive to trochaic stress they should show different patterns of discrimination.

---

[2] The learning algorithm involved sibling-descendant cascade-correlation. Cascade-correlation networks which recruit hidden units as needed have been useful for simulating various developmental phenomena, as compared to back-propagation learning algorithms that adjust weights in static network architectures (Schultz, 2003).

[3] See Curtin and Werker (Ch. 35 this volume) on phonological development for an extensive review of the literature on infant word segmentation.

Nine-month-olds listened longer to words than to part-words in the trochaic condition, but showed the reverse pattern in the condition with iambic stress. The latter result is consistent with having segmented words on the basis of trochaic stress. For these infants (in the iambic condition), trochaically segmented words began with the second stressed syllable of a word (as defined by high transitional probability) and ended with the first syllable of the next word, and thus consisted of part-words (two syllables spanning words). Apparently, the lower transitional probabilities of such combinations were overwritten by the powerful cue of trochaic stress, leading infants to prefer these. Six month olds were oblivious to the stress manipulation regardless of whether they were in the trochaic or iambic stress condition, demonstrating that when both cues are present younger infants will favor statistical information. This study is important for suggesting how infants might bootstrap one kind of information (knowledge of stress patterns) from another (transitional probabilities).

36.4.2 Using information obtained in one linguistic task as input to another level of learning

We have seen how learning of one kind of structure (transitional probabilities) interacts with learning of another kind (stress) at different points in development, suggesting that what learners acquire at one point in developmental time can impact later learning. How do infants fare when required to learn at multiple levels? Saffran and Wilson (2003) examined this question by asking whether twelve month olds would use the output of one process (word segmentation) as the input to another (learning the sequential ordering of words). Linguistic input is not organized such that children are first taught words, then syntax. Rather, children experience language holistically. As such, they must segment words, based on patterns of syllables or sublexical units, as a precursor to learning their sequential ordering. Saffran and Wilson exposed infants to strings such as *datopidubutobadudipa* where *dato*, *pidu*, *buto*, *badu*, and *dipa* were words arranged with constraints on their sequential ordering in strings. Infants were then tested on novel legal and illegal strings. Transitional probabilities between syllables were identical in the two sentence types (1.0 within words and 0.25 between words). Thus, in order to discriminate the sentence types, infants must have segmented words. An optional first syllable was included to prevent detection of grammatical strings based on absolute position of syllables in strings. Infants discriminated legal and illegal strings, suggesting that they were applying what they had learned as input at one level to a subsequent learning process. A question, however, is whether discrimination reflected sensitivity to the ordering of segmented words or to runs of four or more syllables.

36.5 Rudiments of syntax

36.5.1 Learning sequential word ordering

The studies reviewed thus far show that infants rapidly detect statistical structure at the level of segments and syllables, and can form generalizations from this information. The next step is to ask whether infants can track words in strings. Although syntax acquisition ultimately involves the ability to track categories in phrases (e.g. Determiner, Noun) and the hierarchical organization of phrases in strings, it is informative as a first step to investigate learning of sequential structure. Such investigations yield insights into the kinds of sequential dependencies infants can acquire (Gómez, 2002; Gómez and Gerken, 1999; Gómez and Maye, 2005; Saffran, Hauser, et al., 2005).

In one of the earliest investigations of infants' ability to track the orderings of words in sentences, Gómez and Gerken (1999) exposed twelve month olds to a subset of strings produced by one of two finite-state languages (see also Mintz, 1996). Although word order was constrained by these languages, there was still considerable variability in the orderings of words in strings. For instance, note how the position of the word *pel* varies in the strings *pel-tam-rud*, *vot-pel-jic-rud-tam*, and *vot-pel-pel-jic*. Both languages began and ended with the same words and contained the same vocabulary, but differed in the ordering of word pairs. For instance, the transition *tam-jic* found in Language 1 never occurred in Language 2. After brief exposure to a subset of strings in their training language, infants were given a five-minute play break, and then were tested. Infants listened longer to new strings from their training language than to strings from the other language, regardless of which language they heard during training. Although the constraints placed on word ordering were the same during training and test, infants were never tested on the exact strings

---

Table 36.1 Depicti
in Saffran, Hauser,

Predictive language
S → AP + BP + (CP)ª
AP → A + (D)
BP → CP + F
CP → C + (G)

ª Elements in parenthe
ᵇ Must contain at least
present they must occ

encountered durin
that learning was
for particular stri
to novel strings w
patterns.

Saffran, Hauser,
tested twelve montl
sequential structur
ested in knowing wi
track hierarchical s
in phrases with
between words an
contrasted learning
tive dependencies
with learning of on
were non-predictive
ple, sentences in th
tained an A phrase (
(BP) and an optior
themselves consistec
to four elements ea
encies differed critic
non-predictive lang
in the predictive
required A-word a
A- and D-words we
dictive language (th
phrase). Therefore,
the presence of a D
ence of a particular
tionship in the n
highly variable. Th
dictability of words v
ences in learning suc
discriminate legal an
sure to the predicti
exposure to the nonp

It is unclear whe
learned hierarchica
extremely complex
phrase structure inv
word classes. Altho

t

llable was included to pre-
ımmatical strings based on
syllables in strings. Infants
ınd illegal strings, suggest-
applying what they had
one level to a subsequent
A question, however, is
ion reflected sensitivity to
ıented words or to runs of
s.

ıts of syntax

sequential word

thus far show that infants
:al structure at the level of
:s, and can form general-
ɔrmation. The next step is
nts can track words in
tax acquisition ultimately
track categories in phrases
un) and the hierarchical
:s in strings, it is informa-
ɔ investigate learning of
Such investigations yield
; of sequential dependen-
re (Gómez, 2002; Gómez
Gómez and Maye, 2005;
2005).
: investigations of infants'
derings of words in sen-
Gerken (1999) exposed
subset of strings produced
state languages (see also
;h word order was con-
ıages, there was still con-
:he orderings of words in
ıote how the position of
the strings *pel-tam-rud*,
ıd *vot-pel-pel-jic*. Both
ded with the same words
: vocabulary, but differed
d pairs. For instance, the
ıd in Language 1 never
:. After brief exposure to
their training language,
:-minute play break, and
:s listened longer to new
ining language than to
language, regardless of
heard during training.
s placed on word order-
ring training and test,
ed on the exact strings

Table 36.1 Depiction of the two languages used in Saffran, Hauser, et al. (2005)

Predictive language	Non-predictive language
S → AP + BP + (CP)[a]	S → AP + BP
AP → A + (D)	AP → {(A) + (D)}[b]
BP → CP + F	BP → CP + F
CP → C + (G)	CP → {C + (G)}[b]

[a] Elements in parenthesis are optional.

[b] Must contain at least one category, but if both are present they must occur in left–right order.

encountered during training—demonstrating that learning was not confined to memory for particular strings, but rather generalized to novel strings with familiar co-occurrence patterns.

Saffran, Hauser, et al. (2005) have recently tested twelve month olds on yet more complex sequential structure. Saffran et al. were interested in knowing whether infants this age could track hierarchical structure, such as that found in phrases with predictive dependencies between words and phrases. To do this, they contrasted learning of a language with predictive dependencies between words in phrases with learning of one in which the relationships were non-predictive (see Table 36.1). As an example, sentences in the predictive language contained an A phrase (AP) followed by a B phrase (BP) and an optional C phrase (CP). Phrases themselves consisted of categories A–G with two to four elements each. Within-phrase dependencies differed critically in the predictive and the non-predictive language. Whereas the A phrase in the predictive language consisted of a required A-word and an optional D, both A- and D-words were optional in the non-predictive language (the same was true for the C phrase). Therefore, in the predictive language the presence of a D-word guarantees the presence of a particular A-word,whereas the relationship in the non-predictive language is highly variable. These variations in the predictability of words within phrases led to differences in learning such that infants were able to discriminate legal and illegal strings after exposure to the predictive language, but not after exposure to the nonpredictive one.

It is unclear whether infants in this study learned hierarchical phrase structure or an extremely complex finite-state grammar. A phrase structure involves knowledge of abstract word classes. Although there were multiple words in each of the A–G categories, the word classes were very small, and thus the language can be characterized as a finite-state grammar with transitions between words instead of word classes (Altmann et al., 1995). An important control was that legal and illegal test items were equated in terms of co-occurrence statistics; thus learning could not have stemmed solely from bigram statistics between pairs of words. Rather, the coherence in the predictive language appeared to lead to stronger perceptual grouping of words in phrases. But, perceptual grouping does not imply knowledge of abstract word classes. Additionally Gómez and Gerken have been unable to find evidence of tweleve month olds being able to learn syntactic categories in a much simpler paradigm (unpublished research). Thus, the likelihood that infants this age are using syntactic categories (required for a phrase structure grammar) is low. Regardless of whether infants learned a finite-state grammar or phrase structure, the learning required was challenging. Infants were exposed to fifty different strings in a twenty three-minute familiarization period requiring them to track dependent relationships between sixteen unique words. Thus, the findings show that infants are capable of tracking complex sequential structure.

36.5.2 Generalization of sequential word order

Researchers have also investigated how infants might generalize sequential structure. Gómez and Gerken (1999) tested this by familiarizing 12-month-olds with a finite-state grammar in one vocabulary and testing them on strings in entirely new vocabulary (infants heard strings like *fim-sog-fim-fim-tup* and were tested on *vot-pel-pel-jic*).[4] Thus, although constraints on word ordering remained the same between training and test, vocabulary did not. Because test strings were instantiated in new vocabulary, learners could not distinguish the two grammars based on transitional probabilities between remembered word pairs. Infants made the discrimination, suggesting that they had abstracted something about grammatical structure above and beyond pairs of specific words.

---

[4] The strings in this example do not have identical repetition structure because infants were exposed not only to different vocabulary between training and test, but also to different strings. Although different, the strings retained regularities dictated by the artificial language.

Marcus et al. (1999) reported similar findings for younger seven-month-olds exposed to simple ABA versus ABB (*wi-di-wi* vs. *wi-di-di*) patterns. As in Gómez and Gerken (1999), the vocabulary was different between training and test. Infants discriminated strings with the training pattern from those with a different pattern despite the change in vocabulary (e.g *ba-po-ba* vs. *ba-po-po*). Marcus et al. further interpreted these findings as evidence that infants are acquiring algebra-like rules involving the substitution of arbitrary elements in abstract variables. They argued that systems which learn from statistical regularities, such as connectionist architectures, are in principle incapable of such generalization (see also Marcus, 2001). However, generalization based on associative learning has been demonstrated in a number of models (Altmann, 2002a; Altmann and Dienes, 1999; Christiansen and Curtin, 1999; Gassar and Colunga, 1999; Seidenberg and Elman, 1999; Shastri, 1999; Shultz and Bale, 2001).

One can also ask to what degree such findings extend into problems of language acquisition. The infant abstraction abilities documented by Marcus et al. (1999) and Gómez and Gerken (1999) are dependent on learning patterns of repeating and alternating elements, e.g. ABB, ABA, ABCA (Gómez et al., 2000). While repeating elements occur in natural language in the form of reduplication, the patterns in these studies occur over physical stimuli in sequence. For example, recognizing *ba-po-ba* and *ko-ga-ko* as instances of the pattern ABA entails noting that the first and last syllables in sequence are physically identical. But most generalizations in syntax involve operations over variables that are not perceptually bound. Compare the pattern-based representation ABA to the category-based representation Noun Verb Noun. Abstracting ABA from *ba-po-ba* involves noting that the first and third elements in a sequence are physically identical, and thus recognition is perceptually bound. In contrast, the Noun Verb Noun relation holds over abstract categories that do not rely on perceptual identity. *Dogs eat bones* and *John loves running* share the same category-based structure, despite the obvious physical dissimilarities between category members such as *John* and *running*. Given this observation, how might we begin to examine learning involving abstract variables?

36.5.3 Category-based abstraction

The ability to perceive category relationships among words in strings is essential to linguistic productivity. An English speaker must be able to generalize from a novel string like *The pleg mooped* to *Is the pleg mooping?*. Such generalization is tremendously powerful: once a novel word is categorized, children can automatically apply syntactic constraints associated with other words in its category. How do children achieve such generalization?

The role of semantics is thought to feature importantly in linguistic category acquisition (Grimshaw, 1981; Pinker, 1984). Children first identify members of semantic categories based on referential information, and then link these to innate knowledge of syntactic categories and functions. But some have proposed that distributional information, in the form of phonological regularities within words of a class, and co-occurrence relations between classes, might also factor into such learning (Braine, 1987; Gleitman and Wanner, 1982; Morgan and Demuth, 1996; Morgan and Newport, 1981; Redington et al., 1998). These two views may not be wholly incompatible. An infant who uses distributional cues to parse categories in speech eventually has to link these categories with semantic referents. Thus, identification of relevant categories in speech could provide a leg-up for the ultimate task of mapping meaning and form (Gómez and Gerken, 2000; Naigles, 2002).

Category-based abstraction has been studied fairly extensively with older learners, and has focused on how learners acquire relations between grammatical classes (Braine, 1987; Frigo and McDonald, 1998; Mintz, 2002; Smith, 1969; Wilson, 2002). Grammatical classes are given arbitrary labels such as a, X, b, and Y. Words from these classes then combine to form legal phrases. For instance, aX and bY might be legal in a language whereas aY and bX are not. Learners are exposed to most, but not all, aX and bY phrases and then are tested to see if they will discriminate new legal phrases from illegal ones. To give an example, imagine that a-elements correspond to *a* and *the* and b-elements to *will* and *can*. Learners will only be successful at discriminating a new legal phrase (e.g. *a cat*) from an illegal one (*a eat*) if they have learned that a-elements go with nouns (the Xs), but not with verbs (the Ys). As in natural language, the functor-like a and b categories have fewer members than lexical-like Xs and Ys.

Gómez and Lakusta (2004) explored such generalization by asking whether twelve month olds would learn the relationship between specific a and b words and X and Y categories. During training infants were exposed to one of two training languages. One language consisted of

---

aX and bY pairin
pairs. Xs were ins
and Ys were mono
used as a cue for
gories. This featur
cues found in na
for instance, noun
than verbs and als
stress (Kelly, 1992
phrases from their
from the other lan
eralization, all X a
Infants discrimina
sentences after a
suggesting that the
relationships betw
the abstract featu
words (syllable nu
occur in natural
exposed to Englis
tional regularities
verbs and link thes
Identifying a re
words and distrib
word classes is onl
Frigo and McDona
associated a and b
then categorize inc
as *a* and *the*) on th
tion with these cu
gories are formed,
for a phrase they
the fact that *the* ar
to make an infere
not heard (e.g. *a c*
word has an X or Y
How can this hy
cial language? We
some X and Y ele
do not have distin
initially rely on X
words into separa
and b categories a
able to predict a Y
regardless of wheth
cues are present (B
Gerken et al. (2
ing with seventee
created a set of st
lexical stems appea
and -*u* and six ma
the case endings -
these experiments
elements. Additic
Xs and Ys were pr
members. For ins
X words containe

novel string like *The pleg*
g mooping?. Such generaliza-
sly powerful: once a novel
, children can automatically
traints associated with other
y. How do children achieve

ntics is thought to feature
uistic category acquisition
Pinker, 1984). Children
ers of semantic categories
information, and then link
owledge of syntactic cate-
s. But some have proposed
nformation, in the form of
rities within words of a
rrence relations between
factor into such learning
tman and Wanner, 1982;
uth, 1996; Morgan and
ington et al., 1998). These
e wholly incompatible. An
tributional cues to parse
eventually has to link these
ntic referents. Thus, identi-
categories in speech could
the ultimate task of map-
orm (Gómez and Gerken,

straction has been studied
h older learners, and has
arners acquire relations
l classes (Braine, 1987;
1998; Mintz, 2002; Smith,
Grammatical classes are
s such as a, X, b, and Y.
ses then combine to form
ance, aX and bY might be
hereas aY and bX are not.
d to most, but not all,
nd then are tested to see
nate new legal phrases
give an example, imagine
spond to *a* and *the* and
can. Learners will only be
nating a new legal phrase
gal one (*a eat*) if they have
s go with nouns (the Xs),
ne Ys). As in natural lan-
e a and b categories have
xical-like Xs and Ys.
2004) explored such gen-
hether twelve month olds
nship between specific a
nd Y categories. During
exposed to one of two
ne language consisted of

aX and bY pairings, the other of aY and bX pairs. Xs were instantiated as disyllabic words and Ys were monosyllabic. Syllable number was used as a cue for distinguishing X and Y categories. This feature was chosen to mimic similar cues found in natural language. In English, for instance, nouns tend to have more syllables than verbs and also tend to receive first syllable stress (Kelly, 1992). Infants were tested on new phrases from their training language vs. phrases from the other language. In order to assess generalization, all X and Y words were novel at test. Infants discriminated between legal and illegal sentences after a short familiarization period, suggesting that they had become sensitive to the relationships between the a and b elements and the abstract feature differentiating X and Y words (syllable number). Similar learning may occur in natural language, where children exposed to English may pick up on distributional regularities distinguishing nouns and verbs and link these to specific function words.

Identifying a relationship between function words and distributional cues differentiating word classes is only the first step (Braine, 1987; Frigo and McDonald, 1998). After learners have associated a and b words with X/Y cues, they can then categorize individual a or b elements (such as *a* and *the*) on the basis of their joint association with these cues. Once function-word categories are formed, learners can rely on memory for a phrase they have heard (e.g. *the cat*) and the fact that *the* and *a* are in the same category to make an inference about a phrase they have not heard (e.g. *a cat*), regardless of whether the word has an X or Y cue.

How can this hypothesis be tested in an artificial language? We can do this by incorporating some X and Y elements into the language that do not have distinguishing cues. Learners may initially rely on X and Y cues to group function words into separate categories, but once the a and b categories are formed, learners should be able to predict a Y word, given a preceding b, regardless of whether or not distinguishing X/Y cues are present (Braine, 1987).

Gerken et al. (2005) investigated such learning with seventeen-month-old infants. They created a set of stimuli in which six feminine lexical stems appeared with the case endings *-oj* and *-u* and six masculine stems appeared with the case endings *-ya* and *-em*. Case endings in these experiments were equivalent to a and b elements. Additionally, cues distinguishing Xs and Ys were present for a subset of category members. For instance, three of six of the X words contained the derivational suffix *-k* (e.g. *polkoj, polku*) whereas three of the Y words contained the suffix *-tel* (e.g. *zhitelya, zhitelyem*). Infants were first familiarized with a subset of stimuli and were then tested to see if they would attend differentially to novel aX and bY stimuli vs. ungrammatical aY and bX ones even when the distinguishing suffix was absent (e.g. generalizing to *vannoj* and *pisarem* after hearing *vannu* and *pisarya*). The infants were able to do this; thus by seventeen months infants have gone beyond the first step of associating particular case endings with cues distinguishing X and Y category members (particular derivational suffixes) to the second one of categorizing the case endings. Having heard *vannu* they were able to treat *vannoj* equivalently. This finding is important for showing that by seventeen months, infants can form categories and dependencies between them from distributional cues in speech.

The ability to abstract categories from sequential information is a significant milestone in cognitive and language development, not only for what it implies about early abstraction abilities but because of its potentially important contribution to syntactic development. It will be important to investigate the next step of how infants link categories parsed in speech with syntactic categories on the basis of referential information.

36.5.4 **Tracking long-distance relationships in strings**

Another milestone for language learners is learning to track remote dependencies. Research thus far shows that infants are adept learners of adjacent sequential structure. However, many dependencies in language are connected by longer-distance dependencies. Some examples are phonemic segments in words, morphosyntactic dependencies between auxiliaries and inflectional morphemes (e.g. *i̲s̲ quickly runni̲n̲g̲*), and dependencies between nouns and verbs in number and tense agreement (*The boy̲s̲ in the tree a̲r̲e̲ laughing*). How easily do learners track remote dependencies in sequential structure?

Newport and Aslin (2004) investigated this question in the context of a word segmentation task. Adult learners had to track non-adjacent syllables in words in a continuous stream of speech. Words consisted of three syllables (e.g. *ba-du-te, ba-to-te, pi-to-ra*) with predictable relationships between the first and last CVs. There were four possible middle CVs in the word set and five non-adjacent CV pairs, with each of the medial CVs occurring in each of the five pairs. The transitional probability between the first and third CV in a word was 1.0, but the other transitions (adjacent CVs or CVs spanning word boundaries) ranged from .20 to .25. Learning was impossible. Learners tested across a series of experiments were unable to track non-adjacent structure (even after ten days of exposure).

Tracking non-adjacent dependencies in a continuous stream of syllables is an exceedingly difficult task, requiring learners to monitor information over different syllables in variable positions in the sound stream. Noting this, Newport and Aslin suggested that perhaps non-adjacent regularities of the type occurring in natural language would be easier to track. Although words do not have dependencies between non-adjacent syllables, such dependencies do occur between phonemic segments (in Hebrew and Arabic, for instance). When the structure of words was changed so that dependencies occurred between non-adjacent consonant segments, (e.g. *p_ g_ t_*) or non-adjacent vowels, (e.g. *a _ u _ e _*), learners segmented the words, demonstrating they had learned the non-adjacent dependencies.

Given the difficulty of tracking non-adjacent dependencies in continuous speech, how do learners fare at tracking non-adjacent dependencies in segmented speech? Research with natural language shows that by eighteen months infants track non-adjacent dependencies over as many as three intervening morphemes (Santelmann and Juszcyk, 1998), discriminating phrases like *is running* from *can running*. Gómez (2002) replicated the Santelmann and Jusczyk findings with 18-month-olds in an artificial language paradigm, investigating conditions that might be necessary for learning. Infants were exposed to one of two artificial languages. Language 1 sentences followed the patterns aXb or cXd (e.g. *pel-wadim-jic, vot-kicey-rud, vot-wadim-rud*). In Language 2 the relationship between the first and third elements was reversed such that *pel* sentences ended with *rud*, and *vot* sentences ended with *jic* (*pel-wadim-rud, vot-kicey-jic, vot-wadim-jic*). In both languages a and c elements were restricted to initial position, b and d elements to final position, and X elements occurred medially. Additionally, adjacent dependencies were identical in both languages (aX occurred in both languages as did Xd: e.g. compare Language 1: aXb and cXd to Language 2: aXd and cXb). Because the two languages were identical with respect to absolute position of elements and adjacent dependencies, they could only be distinguished by noting the relationship between the non-adjacent first and third words.

Gómez manipulated the size of the pool from which she drew the middle element (set size = 3, 12, or 24) while holding frequency of exposure to particular non-adjacent dependencies constant. This manipulation was meant to explore structure found in natural language, where function morphemes come from small sets and lexical morphemes come from much larger ones. She asked whether high variability in the middle element would lead to better perception of non-adjacent dependencies even though the non-adjacent dependencies were equally frequent in all set-size conditions. Eighteen-month-olds were able to acquire the non-adjacent dependency when the intervening element came from a set of 24 possible words, but not when intervening set size was smaller (2 or 12). This finding was replicated in younger fifteen month olds (Gómez and Maye, 2005) and in adult participants using a slightly more complex grammar (Gómez, 2002). At first glance, it seems paradoxical that variability can aid learning. Indeed, one might assume that high variability would add noise and impede learning. However, high variability in the large set-size condition appears to increase the perceptual salience of the non-adjacent words compared to the middle word, and thus facilitates learning.

Non-adjacent dependency learning is important for its potential role in syntactic category abstraction. Evidence comes from Mintz (2003), who isolated the most frequent non-adjacent dependencies in corpora of child-directed speech (Mintz refers to these as "frequent frames"). The words embedded in frequent non-adjacent dependencies tended to be from the same categories (e.g. noun, verb, preposition, adjective, adverb), raising the possibility that the statistical properties of non-adjacent dependencies might lead learners to form categories of elements occurring in particular non-adjacent frames. Mintz has shown that adults and children categorize words in artificial languages as a function of their co-occurrence patterns within frequent frames, lending support to this hypothesis (Mintz, 2002; 2004).

One question is: what kind of mechanism is necessary for such learning? Will a fundamentally associative mechanism suffice, or is one that is specialized for long-distance dependencies required? Onnis et al. (2005) and Elman (pers. commun.) have shown that a simple recurrent network can detect non-adjacent sequential dependencies under the same conditions as humans. The network develops graded

---

representations in
differences betwe
when the middle e

Another questio
studies succeeded
dependencies wh
Newport and Asli
ence is that learne
the added difficul
continuous stream
were already segm
requirements of
However, putting
ture tested by New
the low-variability
(2002). Newport a
learning when the
ments of the same
by elements of a
nants). This made
more perceptible b
ferentiate these ty
studies had no a pr
connected elemen
made the non-ad
High variability o
facilitate word se
et al., forthcomin
type may promote
cent elements in w
non-adjacent depe
also promote learn

The non-adjacen
Gómez reflects onl
tactic structure, r
research. Although
The boy<u>s</u> in the tre
naled by -s and ar
always explicit. In
ing, children m
between the abstra
and the verb are. In
sible link to natur
need to begin inves
dependencies in se

36.5.5 Learnin
structure

Children are expos
sistency in langu
ungrammatical utt
matical utterances
tures vary in their r
degree to which ve
ing for the past ten
to which feminine

n the non-adjacent first and
ted the size of the pool from
middle element (set size = 3,
lding frequency of exposure
djacent dependencies con-
lation was meant to explore
natural language, where
es come from small sets and
ome from much larger ones.
igh variability in the middle
to better perception of non-
es even though the non-adja-
were equally frequent in all
Eighteen-month-olds were
non-adjacent dependency
ng element came from a set
s, but not when intervening
(2 or 12). This finding was
r fifteen month olds (Gómez
in adult participants using a
ex grammar (Gómez, 2002).
ems paradoxical that vari-
rning. Indeed, one might
ariability would add noise
g. However, high variability
ize condition appears to
al salience of the non-adja-
d to the middle word, and
ing.
endency learning is impor-
l role in syntactic category
e comes from Mintz (2003),
ost frequent non-adjacent
orpora of child-directed
rs to these as "frequent
ds embedded in frequent
dencies tended to be from
(e.g. noun, verb, preposi-
rb), raising the possibility
roperties of non-adjacent
lead learners to form cate-
curring in particular non-
ntz has shown that adults
ize words in artificial lan-
of their co-occurrence pat-
frames, lending support to
tz, 2002; 2004).
hat kind of mechanism is
arning? Will a fundamen-
hanism suffice, or is one
r long-distance dependen-
et al. (2005) and Elman
ve shown that a simple
can detect non-adjacent
ies under the same condi-
network develops graded

representations in hidden units which maintain differences between non-adjacent structures when the middle element is highly variable.

Another question is why learners in the Gómez studies succeeded in acquiring non-adjacent dependencies when those in Experiment 1 of Newport and Aslin (2004) failed. A key difference is that learners in Newport and Aslin faced the added difficulty of segmenting words from continuous streams, whereas the Gómez strings were already segmented. This alone makes the requirements of learning very different. However, putting segmentation aside, the structure tested by Newport and Aslin was similar to the low-variability condition tested by Gómez (2002). Newport and Aslin subsequently found learning when they replaced syllables with segments of the same type (e.g. vowels) separated by elements of a different type (e.g. consonants). This made non-adjacent dependencies more perceptible because learners naturally differentiate these types. Learners in the Gómez studies had no a priori reason to group remotely connected elements. Instead, high variability made the non-adjacent relations perceptible. High variability of a middle element may also facilitate word segmentation (see Monaghan et al., forthcoming). Thus, while similarity of type may promote learning between non-adjacent elements in words, any process making the non-adjacent dependency more detectable may also promote learning.

The non-adjacency learning demonstrated by Gómez reflects only some dependencies in syntactic structure, raising challenges for future research. Although the remote dependency in The boys in the tree are laughing is overtly signaled by -s and are, such dependencies are not always explicit. In The boy in the tree is laughing, children must track the relationship between the abstract number of the word boy and the verb are. In order to make a more plausible link to natural language, researchers will need to begin investigating learning of abstract dependencies in sequential structure.

36.5.5 Learning probabilistic structure

Children are exposed to various kinds of inconsistency in language input—in their own ungrammatical utterances and in the ungrammatical utterances of others. Additionally, structures vary in their regularity—e.g. in English the degree to which verbs take the regular -ed ending for the past tense and in Spanish the extent to which feminine nouns end in -a. There are other instances of inconsistency, such as when deaf children are exposed to non-native signs (Newport, 1999; Ross and Newport, 1996; Singleton and Newport, 2004) or when normal hearing children are exposed to pidgin languages. In all these instances, children must distinguish more probable from less probable structure. They must also generalize beyond the data to which they are exposed. A criticism of learning approaches is that the mechanisms are too rudimentary to detect the most appropriate structures in language, that myriad possibilities are available, and that unless the correct structures are detected, learners face a combinatorial explosion of possibilities. The studies reviewed to date suggest that infants can detect structure occurring with perfect probability (when transitional probabilities are 1.0); but what happens when probabilities are lower? Will probabilities be encoded accurately or will they be changed in some way?

Hudson Kam and Newport (2005a) tested adults and six year olds on an artificial language with probabilistic determiners to see whether there would be differences in how learners of different ages encode such input. They hypothesized that children might be more likely than adults to systematize determiner use, given more limited short-term and working memory resources. Adults were exposed to a complex language containing transitive, intransitive, and negative structures, with frequency of occurrence of determiners varying in different conditions (45 percent, 60 percent, 75 percent, and 100 percent). At test, the adults produced determiners at the same rate at which they had occurred during training. A second experiment contrasted adult and child learners. Determiners occurred 60 percent or 100 percent of the time. Adults again matched the probabilistic structure in their language, whereas children diverged. Approximately 14 percent of children in the "60 percent consistent input" condition developed a systematic rule to always use a determiner. None of the adults adopted this rule. Although the percentage of children showing this pattern was small, learning is necessarily limited in artificial language studies. Real-world examples of children exposed to probabilistic language input show a similar pattern of increased systematic rule use (Singleton and Newport, 2004; Ross and Newport, 1996).

The results raise intriguing questions about the role played by children in the transition from pidgin to creole languages. Why is it that children of pidgin speakers produce more systematic versions of their input language than their parents? One proposal is that children systematize input based on access to universal language rules (e.g. Bickerton, 1981). In contrast, Hudson Kam and Newport propose that the tendency to systematize may stem from children's cognitive limitations (see also Newport, 1990). When adults were put under conditions of increasing cognitive complexity they too showed systematization. Adults who were exposed to a language with one predominant determiner and several non-predominant ones were more likely to over-regularize use of the predominant determiner (Hudson Kam and Newport, 2005b).

How well do infants learn on exposure to probabilistic structure? Gómez and Lakusta (2004) exposed twelve month olds to artificial languages with varying degrees of probabilistic structure. In Condition 100/0 all of the training strings were from the infants' "predominant" training language. In Condition 83/17, approximately 83 percent of the training strings were from the predominant language (the remaining 17 percent of the strings were from the other language). In Condition 67/33, the split between the predominant and non-predominant training languages was 67 percent and 33 percent. Infants in the 100/0 and 83/17 conditions learned equally well, whereas learning diminished in the 67/33 condition, suggesting that infants are able to track regularities in probabilistic input even when the regularities did not occur with perfect probability (as was the case in the 83/17 condition). Learning does need to be based on some minimum degree of regularity, as demonstrated by the fact that infants in the 67/33 condition failed to learn.

36.5.6 Bootstrapping from prior learning

A final question has to do with how learning at one point in time impacts learning at another point, and how one experience builds on another. Lany et al. (2005) investigated this question in the context of learning categories in sequential structure. Adult learners were familiarized with aX and bY strings where X and Y elements were distinguished by different morphological endings (e.g. -ee or -oo). There were two each of the a and b words and six each of the X and Y words; thus a's and b's acted as functor-like elements. Pilot testing showed that generalization to unheard cases occurred after extensive exposure to the language, but not after brief exposure. In the experimental manipulation, learners were given extensive exposure in one vocabulary, followed by brief exposure to strings in new vocabulary. All aspects of vocabulary were new (even the morphological endings), requiring learners to transfer a system of underlying relationships. Learners with prior exposure transferred to the new vocabulary after brief familiarization (in contrast with learners who did not have prior exposure), showing that they were able to generalize the structural relationships from experience. Such learning is relevant for structures in language that are more consistently cued than others. For instance, the morphosyntactic dependencies between determiners and noun endings (diminutives and plurals) are more prevalent than for auxiliaries and inflectional verb endings (Lany et al., 2005). Learners could use a process like that studied by Lany et al. to transfer knowledge of morphosyntax relations from noun phrases to similar structure found in verb phrases.

Experience can also be used to bootstrap more complex learning. After exposure to a language involving a key syntactic relationship (the aX/bY language detailed above), adult learners in Lany et al. were able to detect relationships in a more complex language involving acX and bcY structure. This language was particularly challenging because the intervening c-element required learners to track non-adjacent dependencies between a-and-X and b-and-Y elements. Additionally, the more complex language was instantiated in novel vocabulary, forcing learners to draw on their knowledge of the abstract structural relationships of the aX/bY language. Learners with prior experience were able to generalize to the non-adjacent structure. Language learners who did not receive prior exposure to the more simple language did not generalize. This finding is important for showing how learners might scaffold learning of complex structure from learning of more simple forms. Studies in progress show that twelve-month-olds too are capable of generalizing knowledge of simple adjacent dependencies to more complex non-adjacent structure (Lany and Gómez, forthcoming), despite the fact that infants this age have previously been unable to track non-adjacent structure (Gómez and Maye, 2005).

36.6 Summary

Research on infant learning has yielded a wealth of findings and has raised intriguing questions besides. Infants are able to track sequential information in a number of linguistic domains,

---

including phonota
mentation, and re
Infants are also ab
and morphosyntac
ing that learning
productivity. In
appear to be hig
change after as littl
exposure, even
frequently in their
Thiessen, 2003), su
linguistic input or
ongoing.

While the findir
preliminary. Work
et al., 2003), phon
et al., 2002), and s
1996) tap fundame
domains, yet freque
bilities cannot exp
2004). Gambell and
cacy of transitional
a corpus of child-d
probabilities alone
segmentation (the h
ing words was only
this are useful fo
various forms of
respect to scaling l
the levels required
Interestingly, the l
may still be enoug
into the problem of
23 percent of word
proportion of thes
stress. If infants use
tially segment to le
pattern of their nat
Saffran, 2003), they
tional cue (in the fo
for raising segmenta

Yet another cha
primitive units of le
to phonotactic regu
instrumental in lea
structures in one's
agree that some s
basic. Identifying p
for placing a lower
(Yang, 2004) and va
from investigating
primitive structures

Although work o
of long-distance dep
tive, it is still limited
ing syntax. A questio
more akin to that

xtensive exposure in one
y brief exposure to strings
ıll aspects of vocabulary
morphological endings),
·ansfer a system of under-
earners with prior expo-
ıe new vocabulary after
in contrast with learners
r exposure), showing that
·ralize the structural rela-
nce. Such learning is rele-
language that are more
ı others. For instance, the
:ndencies between deter-
ıdings (diminutives and
'alent than for auxiliaries
ndings (Lany et al., 2005).
process like that studied
ısfer knowledge of mor-
from noun phrases to
l in verb phrases.
o be used to bootstrap
ş. After exposure to a lan-
yntactic relationship (the
ed above), adult learners
to detect relationships in
ıge involving acX and bcY
ge was particularly chal-
intervening c-element
ck non-adjacent depend-
X and b-and-Y elements.
e complex language was
ocabulary, forcing learn-
nowledge of the abstract
s of the aX/bY language.
›erience were able to gen-
cent structure. Language
'eceive prior exposure to
uage did not generalize.
rtant for showing how
ld learning of complex
g of more simple forms.
v that twelve-month-olds
generalizing knowledge
endencies to more com-
cture (Lany and Gómez,
the fact that infants this
been unable to track
(Gómez and Maye, 2005).

ning has yielded a wealth
sed intriguing questions
ble to track sequential
er of linguistic domains,

including phonotactics, phonology, word segmentation, and remote dependency learning. Infants are also able to perform phonological and morphosyntactic generalizations, suggesting that learning may contribute to linguistic productivity. In particular, infant learners appear to be highly malleable. They show change after as little as two to three minutes of exposure, even for structure occurring frequently in their native language (Saffran and Thiessen, 2003), suggesting that the influence of linguistic input on learning is pervasive and ongoing.

While the findings are promising, they are preliminary. Work on phonotactics (Chambers et al., 2003), phonemic categorization (Maye et al., 2002), and segmentation (Saffran et al., 1996) tap fundamentally into problems in these domains, yet frequency and transitional probabilities cannot explain all of learning (Yang, 2004). Gambell and Yang (2003) tested the efficacy of transitional probabilities for segmenting a corpus of child-directed speech. Transitional probabilities alone achieved very low levels of segmentation (the hit rate for correctly identifying words was only 23.3 percent). Analyses like this are useful for assessing the success of various forms of learning, particularly with respect to scaling learning performance up to the levels required for real-life language tasks. Interestingly, the low levels reported by Yang may still be enough provide an initial wedge into the problem of word segmentation. Of the 23 percent of words correctly extracted, a high proportion of these must have had trochaic stress. If infants use the subset of words they initially segment to learn the predominant stress pattern of their native language (Thiessen and Saffran, 2003), they would then have an additional cue (in the form of a trochaic template) for raising segmentation performance.

Yet another challenge is determining the primitive units of learning. Although sensitivity to phonotactic regularities could conceivably be instrumental in learning some of the syllable structures in one's native language, most would agree that some subset of syllables must be basic. Identifying primitive units is important for placing a lower bound on statistical learning (Yang, 2004) and valuable insights should arise from investigating the dynamic between the primitive structures of language and learning.

Although work on generalization and learning of long-distance dependencies has been informative, it is still limited in its implications for learning syntax. A question is whether such learning is more akin to that required for acquisition of phonological patterns. Research is needed to investigate learning of abstract long-distance relations like those involved in inflectional agreement and binding, and also to investigate how learners form hierarchical relations over categories. Although current research has uncovered rudiments of syntax-like learning, it will be a challenge to bridge the gap to real-life syntax acquisition.

Finally, there is still a great deal to discover regarding cognitive processes in learning. Although statistical learning occurs in non-linguistic domains (Fiser and Aslin, 2002a, b; Kirkham et al., 2002; Saffran et al., 1999; Saffran, Reeck, et al., 2005), we still know little about the cognitive processes involved. Is some structure more easily learned than others? Is memory retention better for certain types of structure? How lasting are the effects of learning, and how easily is memory for one form of structure overwritten by another form? Answers to these questions may provide information about the precedence of learning cues in terms of how likely they are to be recruited—an issue relevant for how learners choose among multiple types of structure (Gómez, 2002; 2005). Another question is how infants coordinate statistical information in auditory and visual domains—an issue with implications for how learners link meaning and form (Altmann, 2002b). A final question has to do with whether infants are performing computations—an assumption broadly held in the literature. In the case of word segmentation, infants are thought to compute transitional probabilities over adjacent syllables (Aslin et al., 1998; Saffran, 2003). Infants are also thought to perform computations over non-adjacent structure (Newport and Aslin, 2004). Are infants performing computations, or is discrimination the result of a simpler process? One proposal is that learners engage in the natural process of chunking on encountering sequential information (Perruchet and Vintner, 1998). Such chunks are initially randomly formed, but if they match subsequent sequences of syllables they are strengthened in memory. Otherwise, the brain's responsiveness to them fades away. The chunks encountered most frequently win out and constitute the words that are ultimately segmented, as shown in a model successfully reproducing performance levels of human subjects (Perruchet and Vintner, 1998). This account has the benefit of explaining how infants might segment words without imposing the computational demands of tracking every possible bigram. Additional work will be needed to determine the extent to which the account will hold up; but this demonstration raises a crucial question about whether learners must perform computations over linguistic input.

In summary, although a great deal of work is needed, it is no longer realistic to doubt the extent of the contribution of learning to language development. Important questions revolve around determining the basic operations and sensitivities that factor into language acquisition (whether these arise from general analytic processes or take the form of language-specific rules), the contributions of learning from regularities in input, and the interaction of the two.

Acknowledgements

Writing of this chapter was supported by an NSF CAREER award to R. Gómez (BCS-0238584) and NIH R01 HD42170 to LouAnn Gerken, R. Gómez, and E. Plante. Suzanne Curtin, L. A.. Gerken, and Gareth Gaskell provided helpful comments and suggestions.

References

Altmann, G. (2002a) Learning and development in neural networks: the importance of prior experience. *Cognition*, 85: B43–B50.

Altmann, G. (2002b) Statistical learning in infants. *Proceedings of the National Academy of Sciences*, 99: 15250–1.

Altmann, G., and Dienes, Z. (1999) Rule learning by seven-month-old infants and neural networks. *Science*, 284: 875a.

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

Aslin, R., Saffran, J., and Newport, E. (1998) Computation of conditional probability statistics by 8-month-olds infants. *Psychological Science*, 9: 321–24.

Bates, E., and Elman, J. (1996) Learning rediscovered. *Science*, 274: 1849–50.

Bickerton, D. (1981) *Roots of Language*. Karoma Press, Ann Arbor, MI.

Billman, D. (1989) Systems of correlations in rule and category learning: use of structured input in learning syntactic categories. *Language and Cognitive Processes*, 4: 127–55.

Bortfeld, H., Morgan, J., Golinkoff, R., and Rathbun, K. (2005) *Mommy* and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16: 298–304.

Braine, M. (1987) What is learned in acquiring words classes: a step toward acquisition theory. In B MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 65–87. Erlbaum, Hillsdale, NJ.

Chambers, K. (2004) Phonological development: mechanisms and representations. Ph.D. dissertation, University of Illinois.

Chambers, K., Onishi, K., and Fisher, C. (2003) Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87: B69–B77.

Chomsky, N. (1959) A review of B. F. Skinner's *Verbal Behavior*. *Language*, 35: 26–58.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. MIT Pres, Cambridge, MA.

Chomsky, N. (1981) *Lectures on Government and Binding*. Foris, Dordrecht.

Christiansen, M., and Curtin, S. (1999) The power of statistical learning: no need for algebraic rules. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, pp. 114–19. Erlbaum, Mahwah, NJ.

Cutler, A., and Norris, D. (1988) The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14: 113–21.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996) *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, Cambridge, MA.

Fiser, J., and Aslin, R. (2002a) Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99: 15822–6.

Fiser, J., and Aslin, N. (2002b) Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28: 458–67.

Friederici, A., and Wessels, J. (1993) Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics*, 54: 287–95.

Frigo, L., and McDonald, J. (1998) Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39: 218–45.

Gambell, T., and Yang, C. (2003) Scope and limits of statistical learning in word segmentation. In *Proceedings of the thirty fourth Northeastern Linguistic Society Meeting (NELS)*, pp. 29–30. Stony Brook University, New York.

Gassar, M., and Colunga, E. (1999) Babies, variables, and connectionist models. In M. Hahn and S. Stone (eds), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, p. 794. Erlbaum, Mahwah, NJ.

Gerken, L. A. (2004) Nine-month-olds extract structural principles required for natural language. *Cognition*, 93: B89–B96.

Gerken, L. A. (2006) Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, 98: B67–B74.

Gerken, L. A., Wilson, R., and Lewis, W. (2005) 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language*, 32: 249–68.

Gleitman, L., and Wanner, E. (1982) The state of the state of the art. In E. Wanner and L. Gleitman (eds), *Language Acquisition: The State of the Art*, pp. 3–48. Cambridge University Press, Cambridge.

Gómez, R. (2002) Varia
structure. *Psychologi*
Gómez, R. (2005) Dyna
Y. Munakata and M.
Performance XXI: Pr
Cognitive Developme
Press, Oxford.
Gómez, R., and Gerken
language learning an
Cognitive Sciences, 4:
Gómez, R., and Gerken
learning by one-year
knowledge. *Cognitio*
Gómez, R., Gerken, L. A
basis of transfer in a
and Cognition, 28: 25
Gómez, R., and Lakusta
based category abstr
Developmental Scien
Gómez, R., and Maye, J.
trajectory of non-adj
Infancy, 7: 183–206.
Grimshaw, J. (1981) For
acquisition device. In
The Logical Problem
MIT Press, Cambrid
Hudson Kam, C., and N
unpredictable variati
learners in language
Learning and Develo
Hudson Kam, C., and N
by getting it wrong:
MS, submitted for pu
Hyams, N. (1986) *Lang*
Parameters. Reidel, D
Johnson, E., and Jusczyk
8-month-olds: when
statistics. *Journal of*
Jusczyk, P. (1997) *The D*
Press, Cambridge, M
Jusczyk, P., Friederici, A.
Jusczyk, A. M. (1993)
patterns of native lan
and Language, 32: 40
Jusczyk, P., Houston, D.,
beginnings of word s
infants. *Cognitive Psy*
Jusczyk, P., Luce, P., and
sensitivity to phonota
language. *Journal of*
33: 630–45.
Kelly, M. (1992) Using s
the role of phonology
assignments. *Psycholo*
Kemler Nelson, D., Juscz
A., and Gerken, L. A.
Procedure for testing
Behavior and Develop
Kirkham, N., Slemmer, J.
statistical learning in
general learning mec
83: B35–B42.

ological development:
entations. Ph.D. dissertation,

and Fisher, C. (2003) Infants
arities from brief auditory
7: B69–B77.
iew of B. F. Skinner's *Verbal*
26–58.
ts of the Theory of Syntax. MIT

es on Government and Binding.

n, S. (1999) The power of
ed for algebraic rules. *Proceedings*
l Conference of the Cognitive
9. Erlbaum, Mahwah, NJ.
1988) The role of strong syllables
cal access. *Journal of*
: Human Perception and

n, M., Karmiloff-Smith, A.,
K. (1996) *Rethinking Innateness:*
ve on Development. MIT Press,

a) Statistical learning of new
ons by infants. *Proceedings of the*
nces, 99: 15822–6.
b) Statistical learning of
ructure from visual shape
erimental Psychology: Learning,
28: 458–67.
. (1993) Phonotactic
daries and its use in infant
tion and Psychophysics,

(1998) Properties of
at affect the acquisition of
urnal of Memory and Language,

003) Scope and limits of
d segmentation. In *Proceedings*
astern Linguistic Society
0. Stony Brook University,

(1999) Babies, variables, and
M. Hahn and S. Stone (eds),
First Annual Conference of the
p. 794. Erlbaum, Mahwah, NJ.
onth-olds extract structural
tural language. *Cognition,*

ns, decisions: infant language
neralizations are possible.

l Lewis, W. (2005)
stributional cues to form
al of Child Language,

(1982) The state of the
er and L. Gleitman (eds),
State of the Art, pp. 3–48.
ss, Cambridge.

Gómez, R. (2002) Variability and detection of invariant structure. *Psychological Science*, 13: 431–36.

Gómez, R. (2005) Dynamically guided learning. In Y. Munakata and M. Johnson (eds), *Attention and Performance XXI: Processes of Change in Brain and Cognitive Development*, pp.87–110. Oxford University Press, Oxford.

Gómez, R., and Gerken, L. A. (2000) Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4: 178–86.

Gómez, R., and Gerken, L. A. (1999) Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, 70: 109–35.

Gómez, R., Gerken, L. A., and Schvaneveldt, R. (2000) The basis of transfer in artificial grammar learning. *Memory and Cognition*, 28: 253–63.

Gómez, R., and Lakusta, L. (2004) A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, 7: 567–80.

Gómez, R., and Maye, J. (2005) The developmental trajectory of non-adjacent dependency learning. *Infancy*, 7: 183–206.

Grimshaw, J. (1981) Form, function, and the language acquisition device. In C. Baker and J. McCarthy (eds), *The Logical Problem of Language Acquisition*, pp. 165–82. MIT Press, Cambridge, MA.

Hudson Kam, C., and Newport, E. (2005a) Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development*, 1: 151–95.

Hudson Kam, C., and Newport, E. (2005b) Getting it right by getting it wrong: When learners change languages. MS, submitted for publication.

Hyams, N. (1986) *Language Acquisition and the Theory of Parameters*. Reidel, Dordrecht.

Johnson, E., and Jusczyk, P. (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, 44: 548–67.

Jusczyk, P. (1997) *The Discovery of Spoken Language*. MIT Press, Cambridge, MA.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., and Jusczyk, A. M. (1993) Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32: 402–20.

Jusczyk, P., Houston, D., and Newsome, M. (1999) The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39: 159–207.

Jusczyk, P., Luce, P., and Charles-Luce, J. (1994) Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33: 630–45.

Kelly, M. (1992) Using sound to solve syntactic problems: the role of phonology in grammatical category assignments. *Psychological Review*, 99: 349–64.

Kemler Nelson, D., Jusczyk, P., Mandel, D., Myers, J., Turk, A., and Gerken, L. A. (1995) The Head-turn Preference Procedure for testing auditory perception. *Infant Behavior and Development*, 18: 111–16.

Kirkham, N., Slemmer, J., and Johnson, S. (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83: B35–B42.

Lany, J., Gómez, R., and Gerken, L. A. (2005) The role of prior experience in language acquisition. MS, submitted for publication.

Lany, J., and Gómez, R. (forthcoming) Infants use prior experience to bootstrap from simple to more complex structure. MS in preparation.

Liberman, A., Cooper, F., Shankweiler, D., and Studdert-Kennedy, M. (1967) Perception of speech code. *Psychological Review*, 74: 431–61.

Mackain, K. (1982) Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9: 527–42.

Manzini, R., and Wexler, K. (1987) Parameters, Binding Theory and learnability. *Linguistic Inquiry*, 18: 413–44.

Marcus, G. (2001) *The Algebraic Mind*. MIT Press, Cambridge, MA.

Marcus, G., Vijayan, S., Bandi Rao, S., and Vishton, P. (1999) Rule learning by seven-month-old infants. *Science*, 283: 77–80.

Mattys, S., and Juszcyk, P. (2001) Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27: 644–55.

Maye, J., Werker, J., and Gerken, L. A. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82: B101–11.

Maye, J., Weiss, D., and Aslin, R. (forthcoming) Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*.

Mintz, T. (1996) The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories. Ph.D. dissertation, University of Rochester.

Mintz, T. (2002) Category induction from distributional cues in an artificial language. *Memory and Cognition*, 30: 678–86.

Mintz, T. (2003) Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90: 91–117.

Mintz, T. (2004) Twelve-month-olds categorize novel words using distributional information. Poster presented at the fourteenth Biennial International Conference on Infant Studies, Chicago, May.

Monaghan, P., Onnis, L., Christiansen, M., and Chater, N. (forthcoming) The importance of being variable: learning non-adjacent dependencies in speech processing. MS under review.

Morgan, J., and Demuth, K. (eds) (1996) *Signal to Syntax*. Erlbaum, Mahwah, NJ.

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

Naigles, L. (2002) Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition*, 86: 157–99.

Newport, E. L. (1990) Maturational constraints on language learning. *Cognitive Science*, 14: 11–28.

Newport, E. (1999) Reduced input in the acquisition of signed languages: contributions to the study of creolization. In M. DeGraff (ed.), *Language Creation and Language Change*, pp. 161–78. MIT Press, Cambridge, MA.

Newport, E., and Aslin, R. (2004) Learning at a distance, I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48: 127–62.

Onnis, L., Destrebecqz, A., Christiansen, M., Chater, N., and Cleeremans, A. (2005) Processing non-adjacent dependencies: a graded, associative account. MS, submitted for publication.

Pegg, J., and Werker, J. (1997) Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102: 3742–53.

Perruchet, P., and Vintner, A. (1998) PARSER: a model for word segmentation. *Journal of Memory and Language*, 39: 246–63.

Pinker, S. (1984) *Language Learnability and Language Levelopment*. Harvard University Press, Cambridge, MA.

Redington, M., Chater, N., and Finch, S. (1998) Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*, 22: 425–69.

Roeper, T., and Williams, E. (1987) *Parameter Setting*. Reidel, Dordrecht.

Ross, D., and Newport, E. (1996) The development of language from non-native linguistic input. In A. Stringfellow, D. Cahana-Amitay, E. Hughs, and A. Zukowski (eds), *Proceedings of the twenteeth Annual Boston University Conference on Language Development*, pp. 634–5. Cascadilla Press, Boston, MA.

Saffran, J. (2001) Words in a sea of sounds: the output of statistical learning. *Cognition*, 81: 149–69.

Saffran, J. (2003) Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, 12: 110–14.

Saffran, J., Aslin, R., and Newport, E. (1996) Statistical learning by eight-month-old infants. *Science*, 274: 1926–28.

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Fritz, T., and Cushman, F. (2005) Cross-species differences in the capacity to acquire language: Grammatical pattern learning by human infants and monkeys. MS, submitted for publication.

Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999) Statistical learning of tonal structure by adults and infants. *Cognition*, 70: 27–52.

Saffran, J., Reeck, K., Niehbur, A., and Wilson, D. (2005) Changing the tune: absolute and relative pitch processing by adults and infants. *Developmental Science*, 8: 1–7.

Saffran, J., and Thiessen, E. (2003) Pattern induction by infant language learners. *Developmental Psychology*, 39: 484–94.

Saffran, J., and Wilson, D. (2003) From syllables to syntax: multi-level statistical learning by twelve-month-old infants. *Infancy*, 4: 273–84.

Santelmann, L., and Jusczyk, P. (1998) Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition*, 69: 105–34.

Seidenberg, M., and Elman, J. (1999) Do infants learn grammar with algebra or statistics? (letter) *Science*, 284: 434–35.

Shastri, L. (1999) Infants learning algebraic rules. *Science*, 285: 1673–74.

Shultz, T. (2003) *Computational Developmental Psychology*. MIT Press, Cambridge, MA.

Shultz, T., and Bale, A. (2001) Neural network simulation of infant familiarization to artificial sentences: rule-like behavior without explicit rules and variables. *Infancy*, 2: 501–36.

Shultz, T., and Gerken, L. A. (2005) A model of infant learning of word stress. In *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society*, pp. 2015–20. Erlbaum, Mahwah, NJ.

Singleton, J., and Newport, E. (2004) When learners surpass their models: the acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, 49: 370–407.

Skinner, B. F. (1957) *Verbal Behavior*. Appleton-Century-Crofts, New York.

Smith, K. (1969) Learning co-occurrence restrictions: rule learning or rote learning? *Journal of Verbal Behavior*, 8: 319–21.

Stager, C., and Werker, J. (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388: 381–2.

Swingley, D., and Aslin, R. (2000) Spoken word recognition and lexical representation in very young children. *Cognition*, 76: 147–66.

Thiessen, E., Hill, E., and Saffran, J. (2005) Infant-directed speech facilitates word segmentation. *Infancy*, 7: 53–71.

Thiessen, E., and Saffran, J. (2003) When cues collide: use of stress and statistical cues to word boundaries by 7- to nine-month-old infants. *Developmental Psychology*, 39: 706–16.

Werker, J., Fennell, C., Corcoran, K., and Stager, C. (2002) Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*, 3: 1–30.

Werker, J., and Pegg, J. (1992) Infant speech perception and phonological acquisition. In C. Ferguson, L. Menn, and C. Stoel-Gammon (eds), *Phonological Development: Models, Research, Implications*, pp. 285–311. York Press, Timonium, MD.

Werker, J., and Tees, R. (1984) Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology*, 37: 278–86.

Wexler, K., and Culicover, P. (1980) *Formal Principles of Language Acquisition*. MIT Press, Cambridge, MA.

Wilson, R. (2002) Category induction in second language learning: what artificial grammars can tell us. Ph.D. dissertation, University of Arizona.

Yang, C. (2004) Universal grammar, statistics or both? *TRENDS in Cognitive Science*, 8: 451–6.

CHAP

Wor

Melissa A.

37.1 Introdu

What is a word? I
sounds, something
gle breath? Is it a
picks out some
event, or perhaps a
description, but o
(e.g. *the, a, of*) do n
meaning due to its
ture of a sentence?
roles, but it may n
have meaning for
they can string the
a sound sequence
with clear referent
special attention
when learning new
to learn and recog
such cues (Werker

In this chapter,
word learning must
of a word. Word lea
problem precisely b
plex collections of p
It is a linguistic unit
cal, and phonolog
social convention. It
that refers to a mea
dren's learning of w
as *dog, water*, and *M*
capacities that inter

The multi-facete
challenge in develo
to further developm
to piece out the crit
this knowledge the
future learning. As H
and their colleagues
requires a "coalitio
integrated during de