ANNUAL REVIEWS

*Annual Review of Linguistics*

# Artificial Language Learning in Children

Jennifer Culbertson[1] and Kathryn Schuler[2]

[1]Centre for Language and Evolution, University of Edinburgh, Edinburgh EH8 9AD, United Kingdom; email: jennifer.culbertson@ed.ac.uk

[2]Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: kschuler@sas.upenn.edu

Annu. Rev. Linguist. 2019. 5:353–73

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-011718-012329



## Keywords

language acquisition, artificial language learning, generalization, linguistic typology

## Abstract

Artificial language learning methods—in which learners are taught miniature constructed languages in a controlled laboratory setting—have become a valuable experimental tool for research on language development. These methods offer a complement to natural language acquisition data, allowing researchers to control both the input to learning and the learning environment. A large proportion of artificial language learning studies has aimed to understand the mechanisms of learning in infants. This review focuses instead on investigations into the nature of early linguistic representations and how they are influenced by both the structure of the input and the cognitive features of the learner. Looking not only at young infants but also at children beyond infancy, we discuss evidence for early abstraction, conditions on generalization, the acquisition of grammatical categories and dependencies, and recent work connecting the cognitive biases of learners to language typology. We end by outlining important areas for future research.

## 1. ARTIFICIAL LANGUAGE LEARNING AS A COMPLEMENT TO NATURAL LANGUAGE EVIDENCE

Understanding how children's knowledge of language develops can shed light on how the capacity for language evolved in our species, how language is shaped by our cognitive and social environment, and the extent to which it requires a specialized system. Much of the research in this area uses natural language corpora and experimental methods to test infants' and children's ability to perceive, comprehend, and produce their language. Indeed, researchers have developed and tested key ideas and hypotheses—from perceptual reorganization (e.g., Werker & Tees 1984) to the poverty of the stimulus (e.g., Pullum & Scholz 2002) to statistical learning (e.g., Yang 2004, Peperkamp et al. 2006)—on the basis of data from natural language experiments and corpora. However, there are limitations inherent in natural language data. Most fundamentally, natural languages are complex; isolating one part of a linguistic system from the rest of the language is difficult, if not impossible, and children do not learn each part of their language in a vacuum.

These intricate connections pose a challenge if we want to make claims about how a particular sound, pattern, or structure is (or can be) learned. For example, some researchers have argued that children must come to the task of learning with certain assumptions (e.g., that language is hierarchically structured or that certain semantic or syntactic categories exist). In some cases, this is based on the argument that the input does not provide enough evidence for children to learn these things (e.g., Crain & Nakayama 1987). However, we typically do not have reliable information about the specific input any particular child has received. Further, the influence of the larger linguistic context can obscure exactly what information is available to support learning: There may be alternative sources of evidence that lead a learner toward (or away from) a specific hypothesis, and we cannot control whether or how frequently children are exposed to these. Natural language data also present difficulties for testing hypotheses about the nature of children's linguistic representations. For example, if children are able to understand or use a particular structure at a given age, this does not necessarily mean that their representation of that structure is adult-like. As we discuss in detail below, children's early representations may be just as abstract as adults', or they may be less abstract (e.g., reflecting memorized chunks or strings memorized from their input).

Finally, many linguistic theories have posited connections between learning and language typology (see, e.g., Chomsky & Hale 1968, Baker 2001, Adger 2003, Hayes et al. 2004). Certain types of structures or patterns appear to be underrepresented or unattested in the world's languages, which may reflect constraints or biases in learning. However, this is challenging—and in some cases impossible—to verify on the basis of natural language data. Most obviously, if a particular pattern is unattested, we cannot ask whether children are able to acquire it in a natural language. Relatedly, showing that different patterns are not equally learnable cannot easily be done by comparing ease of acquisition across natural languages, because this will be confounded by the many other differences among them.

All the issues identified above can be tackled by providing independent evidence from artificial language learning[1] (ALL) experiments, which allow us to carefully construct "languages" that differ minimally from each other and to control precisely the input learners are exposed to. These methods make it possible to test how strongly a learner's knowledge is tied to precisely those exemplars (e.g., specific words, strings) that are present in the input. How abstract representations are at any given time makes predictions about how learners should generalize to new data. This can be straightforwardly tested when the input is determined by the experimenter and the test

[1] Sometimes also called artificial "grammar" learning.

data are explicitly constructed to test alternative hypotheses. We can also instantiate any logically possible linguistic structure or pattern in an artificial language, even those that are not attested in natural languages. This allows researchers to test hypothesized constraints on learning, which would not be possible (either practically or in principle) using natural language data. Notably, ALL can and has been used with learners of all ages, from infants to adults. Although experiments with adults often provide the initial tests of a hypothesis, the primary learners of language are not adults, and adults differ from children in terms of both cognitive capacity and linguistic and world knowledge. This article therefore focuses on infant and child learners, mentioning closely related work on adults when relevant.



## 2. IN THE BEGINNING

Among the earliest uses of ALL were the classic experiments by Reber (1967, 1969) and Braine (1963) that examine whether humans could gradually become sensitive to structural patterns in language-like stimuli without explicit instruction or conscious knowledge. Braine (1963) taught 9- to 10-year-old children sequences of nonsense words. Unbeknownst to the children, the words were separated into two classes, A and B, and followed a single rule: All phrases in the language comprised an A word followed by a B word. Children quickly learned this underlying rule, and more importantly, they generalized it to new combinations of words. These methods were subsequently applied to investigate learning and generalization of more complex structures, generally with adults. For example, a number of early studies tested whether specific features of natural language, such as cues to constituent structure and agreement, facilitated learning (e.g., Morgan & Newport 1981; Morgan et al. 1987, 1989; Reber 1989; Brooks et al. 1993).

In the 1990s, a parallel strand of research began using ALL to investigate how infants learn to pick out individual words in continuous speech (Morgan & Saffran 1995, Saffran et al. 1996, Aslin et al. 1998). These studies addressed the mechanisms of learning, in particular whether statistical (or distributional) learning underlies early acquisition. Statistical learning refers to the idea that learners track, store, and use frequency, position, and co-occurrence information to discover underlying structure in the input. This structure most famously includes word boundaries, but also phonotactic restrictions (information about what sounds can occur and where), syntactic categories and constituent structure (information about word-level and phrase-level groupings), morphology (local and nonlocal relationships among word forms), and word order. There is now general agreement in the field of language acquisition that statistical learning of some kind is a major mechanism by which children acquire language, and many excellent reviews on this topic exist (often covering both infant and adult work: e.g., Gómez & Gerken 2000, Saffran 2003, Romberg & Saffran 2010, Aslin & Newport 2012, Wonnacott 2013, Erickson & Thiessen 2015). Although there has been skepticism about the naturalness of both the constructed language-like systems and the learning environment, some recent articles suggest a reasonably close link between artificial and natural language learning (e.g., Gómez & Gerken 2000, Petersson et al. 2004, de Vries et al. 2008, Friederici et al. 2011, Misyak & Christiansen 2012, Ettlinger et al. 2016).

Despite early studies with older children, most ALL studies target either infants or adults. Here we focus not on statistical learning or other mechanisms of learning per se, but rather on the nature of early representations and on how these are affected by features of the input on the one hand and cognitive features of the learner on the other. We say early representations because we are talking about the first stages of learning, when only limited input has been received in the new language-like system. However, the issues we focus on have led to studies with both infants and older children. Therefore, this review is unique in covering the use of ALL with children beyond infancy.

# 3. THE NATURE OF EARLY REPRESENTATIONS

## 3.1. How Abstract?

The nature of learners' early representations is a contentious issue in language acquisition research because it relates to a larger debate around our language learning capacity. On one hand, if after a limited amount of exposure learners are predominantly sensitive to surface properties of the utterances they have heard, then early representations of natural language may predominantly reflect memory of previously heard items. Researchers who argue that early representations are item based often reason that general-purpose learning mechanisms are likely sufficient for language acquisition (e.g., Tomasello 2000). On the other hand, if after limited exposure learners are sensitive to higher-level properties of what they have heard—structural features such as what kinds of words precede others, or how different types of words are grouped—then even very early representations of natural language may reflect this level of abstraction. Researchers who argue that early knowledge is abstract often advocate the position that language acquisition reflects a species-specific predisposition, with potentially specialized language learning mechanisms (e.g., Lightfoot 1999). It is therefore theoretically important to understand the nature of early representations. However, using evidence of early abstraction (or not) in natural language acquisition to adjudicate between theories has proven problematic, particularly in the absence of sophisticated models of learning.

ALL studies have allowed researchers to assess early stages of learning in very young (even prelinguistic) infants under controlled conditions. Building on the work of Braine (1963), Mintz (1996) used the Head-Turn Preference Procedure[2] to train 9-month-old infants on sequences of syllables generated by a miniature artificial language. Infants subsequently listened longer to strings they had heard during training and strings in unfamiliar orders. This suggests that young infants can learn sequential information following a very brief exposure; however, it is unclear whether infants remembered the specific strings they heard during training or whether they represented the structural properties of the input (e.g., words of class A precede words of class B). To test whether infants represent structure-based information in this context, Gómez & Gerken (1999) trained 1-year-olds on strings generated from a finite-state grammar (as in **Figure 1**, following Reber 1967) using the Head-Turn Preference Procedure. Crucially, infants were trained on one set of sequences but were tested on novel sequences conforming to the grammar or not. Infants listened longer to new strings that conformed to the grammar, suggesting that their knowledge was sufficiently abstract to allow generalization beyond the stimuli they were trained on. This was replicated even when infants were trained on a grammar using one set of vocabulary items and tested using an entirely different set. The knowledge infants have after only a very short exposure to new language-like stimuli (several minutes in this case) thus involves a degree of abstraction that goes beyond simply memorizing strings of words.

The ability of very young infants to extract abstract, structure-based information from simple sequences is further supported by Marcus et al. (1999), who showed that 7-month-old infants could learn AAB or ABA patterns containing arbitrary elements. Infants were made to listen to

[2]In the Head-Turn Preference Procedure (Kemler Nelson et al. 1995), infants are seated on a caregiver's lab in front of a setup typically consisting of a central light and two peripheral speakers. The infant is oriented to the central light at the beginning of each trial. A stimulus is then played from one of the speakers for as long as the infant orients toward it. The trial ends when the infant looks away (e.g., for more than two seconds). The experimenter records listening (or orientation) times for each trial. Often, there is a short break between exposure—when infants are trained on a set of stimuli—and testing. In this paradigm, learning is demonstrated by infants who distinguish between the stimuli they were trained on and novel stimuli. Sometime this means longer orientation (or listening) to familiar items, but more often it is longer listening to novel items (for discussions of direction of effects see, e.g., Houston-Price & Nakai 2004, Aslin 2007).

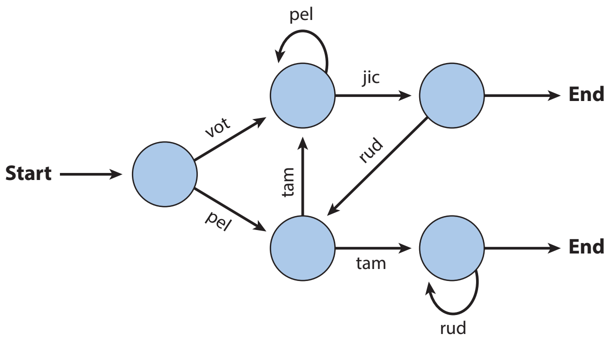


**Figure 1**

Finite-state grammar generating strings such as *vot pel jic* in an artificial language learning study by Gómez & Gerken (1999). The study tests whether young infants acquire structure-based information following brief exposure to strings generated from the grammar.

strings conforming to either AAB or ABA (e.g., *le le di* or *le di le*) until they were habituated. Subsequently, they heard totally new strings conforming to that pattern or not.[3] Infants listened longer to the new, nonconforming strings. Because the test strings did not overlap in content with the training string, infants must have picked up on the structural regularity. Interestingly, this is not always the case: Infants must have sufficient evidence of the generality of the pattern. If the input contains AAB strings that all end in the syllable *di*, for example, infants learn a more restrictive generalization (something like AA*di*; Gerken 2006). Abstract pattern learning of this kind can also be used to learn naturalistic phonological rules. Gerken (2004) found that 9-month-old infants exposed to strings conforming to a set of stress assignment rules (e.g., stress heavy syllables, do not stress adjacent syllables) learned the stress patterns they were exposed to and inferred that a new pattern was ruled out on the basis of the input.

## 3.2. Knowing When to Generalize

The experiments discussed above show that infants readily generalize on the basis of specific strings in their input. They also suggest that representations of a generalization can be more or less abstract depending on the input (e.g., AAB versus AA*di* in Gerken 2006). Precisely when a learner should generalize widely, and how exceptions can still be acquired, is a pervasive issue for acquisition known as Baker's paradox (Baker 1979). For example, infants acquiring English must learn to generalize the regular plural morpheme *-s* without overapplying it (e.g., to irregulars such as *children*). The same goes for learning verbal inflection (e.g., past tense *-ed*) and more complex constructions such as the dative alternation (e.g., the double-object construction *give a book to the library* or *give the library a book*, which is not available for verbs such as *donate*). In all these cases, there is evidence that children first overgeneralize, but what specific features of the input affect the likelihood of generalization remains to be fully understood.

Several recent ALL experiments have explored this. Wonnacott (2011) taught 5- to 7-year-olds a language in which nouns were followed by one of two particles (see Wonnacott et al. 2008

[3]Habituation paradigms are very similar to Head-Turn Preference. In this case, the training stimuli are played until the infant is habituated (as indicated by failure to attend to the stimulus); then a new type of stimulus is introduced (often without any break) to test whether infants distinguish it from the training stimuli. If they do distinguish the new stimulus, it should grab their attention; if they do not, they will continue to show habituation.

**Day 1**

| Lexicalist | Generalist |
|---|---|
| *noun*$_1$ *particle*$_1$<br>*noun*$_2$ *particle*$_1$<br>*noun*$_3$ *particle*$_1$<br>*noun*$_4$ *particle*$_2$ | *noun*$_{\{1,2,3,4\}}$ *particle*$_1$<br>*noun*$_{\{1,2,3,4\}}$ *particle*$_2$ |

**Day 2**

| Lexicalist | Generalist |
|---|---|
| *noun*$_5$ *particle*$_1$<br>*noun*$_6$ *particle*$_2$ | *noun*$_5$ *particle*$_1$<br>*noun*$_6$ *particle*$_2$ |

**Figure 2**

Input for Lexicalist and Generalist languages across two days of training in an artificial language learning study by Wonnacott (2011). This study compares learners' patterns of generalization based on two input conditions: Generalist (in which nouns freely co-occur with two particles) and Lexicalist (in which a given noun co-occurs with only a single particle).

for a similar experiment with adult learners). In the Lexicalist condition, three nouns occurred exclusively with one particle and one noun occurred exclusively with the other. In other words, noun–particle co-occurrence was lexically determined. In the Generalist condition, all four nouns occurred with both particles (though one was three times more frequent). In both conditions, children were trained during one session and brought back in for a second session. In this session, they were first tested on the nouns they had learned previously; then they were trained on a small number of sentences with new nouns occurring exclusively with either one particle or the other. The input over both days for each condition is summarized in **Figure 2**.

Children in both conditions learned the patterns of particle usage they were trained on. If a noun appeared with only one particle in the input, children reproduced this pairing. If nouns appeared with both particles, children reproduced them with both. Interestingly, this was extended to new nouns in the second session; these nouns occurred with only a single particle, but children in the Generalist condition were more likely to reproduce them with both particles. In other words, children were influenced by the higher-level properties of their input on the first session: If the language they were learning was one in which particle usage was lexically determined, they extended this property to the new nouns learned. However, if the language they were learning allowed free alternation, they assumed new nouns would alternate as well. Unsurprisingly, children in both conditions were also sensitive to the frequency of the particles. At the end of day two, children were asked to form sentences with nouns they had never heard used with any particles. In both conditions, children were more likely to use the particle that occurred more frequently in their input. In a follow-up experiment, Wonnacott et al. (2017) showed that the type of skewed input child learners received in this experiment (i.e., one particle is very common, whereas the other is rare) actually facilitates learning of lexically conditioned variation like this.

The well-known wug test (Berko 1958) supports the idea that young children readily generalize regular forms to novel words. In this experiment, children were given the base form of a novel noun and prompted to provide its inflected form (e.g., *This is a wug. Now there are two of them. There are two* ___). Berko tested children on a number of English rules, including plural, possessive, and progressive and past tenses. In each case, children correctly extended the regular form to novel words, suggesting they had formed a productive generalization. Importantly, when prompted with novel words analogous to irregular forms (e.g., *This man knows how to gling. He is glinging. Yesterday he* ____), children used the regular form rather than generalizing an irregular pattern (e.g., *glinged*, not *glung*).

Schuler et al. (2016) used a similar paradigm to explore precisely the conditions under which children generalize a pattern. In particular, they tested the Tolerance Principle (Yang 2005), which describes the precise number of exceptions a learner will tolerate while still forming a productive generalization: The number of exceptions cannot exceed the number of words in the relevant category $N$ divided by the natural log of $N$. Schuler et al. (2016) exposed 5- to 8-year-old children to a nine-noun artificial language in which some nouns took a regular plural form and others were exceptions to this rule. In this case, the Tolerance Principle states that generalization should occur only when the number of exceptions does not exceed four [i.e., 9/log(9)]. As predicted, when prompted to provide the plural form of novel items, children exposed to a language with five regular forms and four exceptions generalized, applying the regular form in nearly 100% of cases. However, when the number of exceptions exceeded the threshold, in this case three regular forms and six exceptions, children failed to make this generalization.

### 3.3. Evidence for Feature-Based Representations

The experiments outlined above suggest that learners likely form representations of linguistic knowledge at multiple levels. Evidence that this involves abstraction early on comes from measures of how learners generalize, making inferences or predictions about new sequences or strings they have not heard. A related question for early learning of phonology is whether infants base these abstractions on something like phonological features—the hypothesized basic units of phonological structure grouping sounds into classes. For example, if a learner hears a set of words in which *p*, *t*, and *k* appear only as the first sound, they can in principle learn a restriction on those individual segments in that position or a restriction on the natural class of voiceless stops. Most theories of phonology postulate feature-based representations, which may reflect an innate set of building blocks or may emerge from infants' early experience with the linguistic input in their environment. In either case, features are potentially valuable: They can help infants discover contrasting phonemes in their language and group together sounds that behave similarly, as in the example above.

It is well established that infants already know a great deal about the phonological structure of their native language before their first birthday (e.g., Werker & Tees 1984; Kuhl et al. 1992; Jusczyk et al. 1993, 1994). In a series of ALL experiments, Saffran & Thiessen (2003) investigated whether 9-month-old infants could rapidly learn a new phonotactic pattern, and if so, whether infants' generalizations were feature based. In one condition, infants were taught a pattern that could be represented on the basis of a single feature: [*p*,*t*,*k*] (voiceless stops) were only ever in syllable onsets, whereas [*b*,*d*,*g*] (voiced stops) were only ever in coda position. In another condition, infants were taught a pattern that could not be represented with a single feature: [*p*,*d*,*k*] were only ever in syllable onsets, whereas [*b*,*t*,*g*] were only ever in coda position. If infants do indeed employ feature-based representations, we should see a difference in learning across these conditions. Infants in the first condition only need to learn that syllable structure is sensitive to voicing, whereas infants in the second condition must learn a more complex constraint picking out an arbitrary set of segments. Using the Head-Turn Preference Procedure, infants were trained on nonsense words following one of these patterns and then heard new exemplars that either followed the training pattern or not. Saffran & Thiessen (2003) found that after training, only infants in the simpler, single-feature condition showed evidence of learning.

Similarly, Cristià & Seidl (2008) trained 7-month-olds on nonsense words in which the first consonant was from a restricted set. In the natural class condition, the onset consonant was always a nasal or oral stop (i.e., [*m*,*n*,*t*,*g*], captured by the single feature [−continuant]). In the arbitrary set condition, the onset consonant was either a nasal or a fricative (i.e., [*m*,*n*,*f*,*z*]). Again, only infants in the natural class condition showed evidence of learning (as before, generalized to new exemplars). In this case, the phonetic similarity of the sounds in the natural class is relatively low (compared with the natural classes used in Saffran & Thiessen 2003), suggesting that the use of abstract phonological features rather than perception of lower-level phonetic similarity is driving these results.

## 4. THE BUILDING BLOCKS OF SYNTAX: CATEGORIES, CONSTITUENTS, AND DEPENDENCIES

### 4.1. Syntactic Categories

Parallel to phonological learning, acquiring syntax involves learning about groups of elements that behave similarly—in this case, syntactic categories such as nouns and verbs. In phonology, sounds are hypothesized to be grouped by learners on the basis of phonetic and distributional similarity. Syntactic categories appear to be similarly built up on the basis of distributional properties such as position, co-occurrence, and semantic similarity. Children's ability to form these kinds of categories has been widely investigated using ALL, revealing both remarkable ability and apparent constraints on learning.

Early work used ALL to investigate whether children can acquire syntactic categories from distributional information alone. Whereas many natural language categories present additional cues (e.g., shared phonological or semantic features), others seem at first glance to be largely arbitrary, the prime example being grammatical gender. Braine et al. (1990) found that children and adults failed to learn two categories of artificial nouns when the only cues to category membership were distributional—for example, each category occurred with a different form of a grammatical marker. However, adult learners succeeded in learning the categories when half of the nouns in each category were semantically similar (stereotypically masculine versus stereotypically feminine occupations; Braine 1987). Whereas Braine et al. (1990) argued that semantic cues were crucial, other work suggests that many kinds of correlated cues can facilitate category learning. Brooks et al. (1993) showed that 9- and 10-year-old children were able to learn artificial categories like those described by Braine (1987), when words shared a phonological feature—a common stem-final vowel.

More recent work has demonstrated that even young toddlers can acquire categories from distributional properties if there are correlated cues to the underlying category structure. Gerken et al. (2005) exposed 17-month-olds to a pseudoartificial language constructed using a subset of the Russian gender paradigm. In the critical experiment, half of the children heard a language in which the only cue to the gender categories was case marking, and the other half heard a double-marked language in which (along with different case markers) each class had a subset of nouns with phonologically similar stems. Using the Head-Turn Preference Procedure, Gerken et al. (2005) demonstrated that only children in the double-marked condition were able to acquire the gender paradigm; children in the single-cue condition failed to distinguish the two classes. In summary, evidence from ALL studies suggests that children's acquisition of categories depends on the existence of rich and (at least partially) overlapping cues in the input. These cues may be distributional, semantic, phonological, or some combination thereof, so long as correlated cues to the category structure are present. Indeed, recent research has shown that both adults (Reeder et al. 2013) and children (K. Schuler, K. Lukens, P. Reeder, E.L. Newport & Aslin, manuscript submitted) can acquire categories from distributional properties alone if the cues are sufficiently rich. It is worth noting that a similar view of the role of correlated cues has emerged from the literature on statistical learning and word segmentation. Infants as young as 6 months of age can use statistical information from adjacent syllables (i.e., which syllable-to-syllable transitions occur most commonly) to pick out candidate words. However, this information must be supported by other cues, such as prosodic changes on word and phrase boundaries (e.g., Johnson & Jusczyk 2001, Johnson & Tyler 2010).

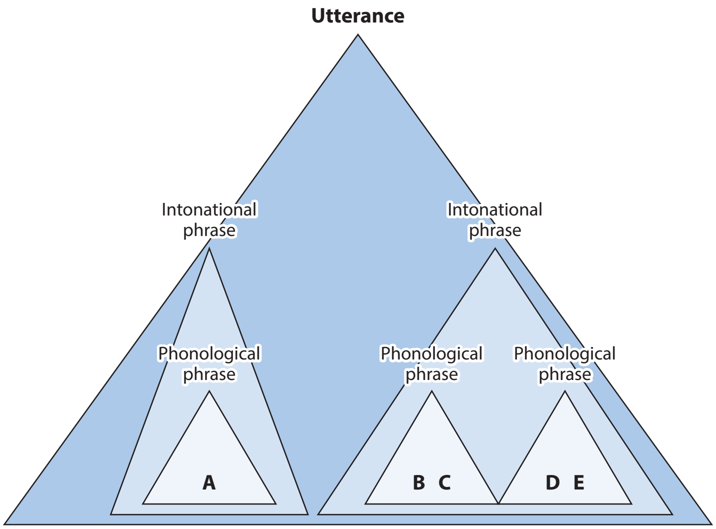


**Figure 3**

Hierarchical structure in the word sequences used by Hawthorne et al. (2016) in their study investigating artificial language learning of hierarchical structure. The utterances included two intonational phrases, one containing a single phonological phrase and the other containing two embedded phonological phrases. A, B, C, D, and E represent five classes of nonce words occurring in the strings learners were exposed to.

## 4.2. Constituents and Hierarchical Structure

Like phonology, syntax involves the acquisition not only of categories but also of structural relations or groupings among elements, such as phrases or constituents. These representations are likely learned concurrently with early syntactic categories. Importantly, syntactic groupings or constituents can be embedded recursively: One constituent can be composed of other subconstituents, which can themselves be made up of subconstituents. A number of studies have suggested that key features of language facilitate or encourage infants to uncover this hierarchical structure. One such property is prosody, the rhythmic structure of language. ALL experiments with adults have shown that prosodic cues can help adults pick up constituents and hierarchical structures (Morgan et al. 1987, Langus et al. 2012).

Hawthorne et al. (2016) investigated the role of prosodic cues to structure using an ALL experiment with 20-month-old toddlers. The input was sequences of words (comprising four classes: A, B, C, and D) that were prosodically grouped as in **Figure 3**. This prosodic structure is akin to what an English speaker would use in a sentence such as *Thankfully, the girl likes cheese*. The structure is hierarchical, containing two main intonational phrases, with the second composed of two embedded prosodic constituents—phonological phrases—parallel to the subject *the girl* and the verb phrase *likes cheese*. The prosodic cues were a pause separating A elements from the longer intonational phrase, and continuation prosody between the two embedded constituents [B C] and [D E].

Infants were trained on these sequences using the Head-Turn Preference Procedure, and then tested on subsequences that either formed part of a prosodic constituent (e.g., [[B C] [D E]] or [B C]) or straddled a prosodic boundary (e.g., A B C D or C D). Toddlers listened longer to items that straddled a prosodic boundary, suggesting that they were able to pick up on the structure of the training sequences. Interestingly, toddlers in another condition, trained on a structure in which the external short intonational phrase was final (as in *The girl liked cheese, thankfully*), did not learn in this task. These findings are intriguing but preliminary, and they call for additional work to explore the conditions under which higher-level hierarchical structure can be learned. We know that sensitivity to syntactic structure, and in particular to nonlocal relations between words in natural language, begins to develop around 18 months of age (Gómez 2002). Prior to that, abstract structure is local in nature, applying to adjacent elements such as segments, syllables, and later words rather than nonadjacent elements. Although infants' early knowledge may be abstract, levels of structure are built up gradually.

## 4.3 Syntactic Dependencies

Phrases or constituents represent one key way in which syntactic elements relate to one another. Prosody is just one way in which these groups can be signaled; another is by formal dependencies among elements, for example, agreement between a noun and a preceding determiner (e.g., in gender or number). These dependencies occur between adjacent elements (local dependencies) or span much longer distances, across phrases and with other elements intervening (nonadjacent dependencies). A number of studies have used ALL to determine whether and under what contexts children can acquire nonadjacent dependencies from distributional information. Several studies suggest that the features of the input may signal such dependencies, facilitating their acquisition. Gómez (2002) exposed adults and 18-month-olds to an artificial language of AXB phrases in which A and B covaried, forming a long-distance dependency similar to the English *that boy-Ø* versus *these boy-s*. Both adults and infants learned the A__B dependency, but only when the number of items in the intervening X category was sufficiently large. If there were only 12 different X elements, adults and infants failed to acquire the dependency; when there were 24 X elements, learners of all ages succeeded. Gómez reasoned that the high variability in the intervening X category made the adjacent dependencies (between AX and XB) less predictive, providing a distributional cue that the relevant dependency is the nonadjacent one.

Additional surface cues to a dependency can also facilitate learning. Newport & Aslin (2004) found that adults could acquire a dependency without a large set of intervening X elements when the AXB pattern was composed of vowel–consonant–vowel or consonant–vowel–consonant segments (e.g., *p__g*), but not when it was composed of syllables (e.g., *pi__ge*). They suggest this may be due to Gestalt principles—primitive principles of perceptual grouping—that guide learners to acquire dependencies between nonadjacent elements that are similar to each other but different from the elements that separate them. If consonants are distinct from vowels in this way, then at some level the dependency is local. Interestingly, the difficult-to-acquire nonadjacent syllable dependencies become easy for adults to learn if the A and B syllables are phonologically similar (Onnis et al. 2005). This suggests that, as with category acquisition, learning of nonlocal syntactic dependencies is facilitated by the presence of correlated cues.

Whereas the input can facilitate and constrain the acquisition of syntactic dependencies, ALL studies have revealed that properties of the learner also constrain what is learned. Gómez & Maye (2005) exposed 12-, 15-, and 18-month-olds to the same AXB language used by Gómez (2002) and found that only the 15- and 18-month-olds learned the nonadjacent dependency. The dependency was not acquired by 12-month-olds, even when they were exposed to the language for a much longer period. Gómez & Maye (2005) suggested that younger babies' limited working memory rendered them unable to acquire the nonadjacent dependencies regardless of the composition of the X category: By the time they processed B, they had already forgotten what A was. However, Lany & Gómez (2008) showed that infants younger than 15 months of age can acquire nonadjacent dependencies if they are first exposed to A and B elements in an adjacent dependency. Local dependencies are thought to be easily acquired by 12 months of age (Gómez & Lakusta 2004). This again suggests that, at least in some cases, more complex linguistic structures are built up from simpler forms (e.g., Newport 1990).

ALL studies in this domain mirror what is known about the developmental trajectory of nonadjacent dependency learning in natural language. Santelmann & Jusczyk (1998) showed that English-acquiring 18-month-olds, but not younger infants, are sensitive to the nonadjacent dependency between *is* and *-ing*. They can distinguish grammatical *the dog is barking at the moon* from ungrammatical *the dog can barking at the moon*, even if some additional material intervenes (*the dog is loudly barking at the moon*). However, if the distance between dependent elements becomes too large (*the dog is quite loudly barking at the moon*), 18-month-olds begin to fail at this task. However, as we saw in the ALL studies, the constraints on learning are more complex than simply the distance between dependencies. Hohle et al. (2006) found that children acquiring German are sensitive to very long distance dependencies if the intervening elements are easier to process: a noun phrase rather than a string of adverbs, for example. Like Newport & Aslin (2004), they suggest that children can group or "chunk" the noun phrase together more easily than a string of adverbs, effectively rendering the distance smaller.

## 5. HOW LEARNING IS CONSTRAINED BY COGNITION

Several studies reviewed above point to the role of the developing cognitive system in explaining certain features of early linguistic representations. For example, ALL experiments on generalization suggest that children may be very sensitive to the number of exceptions when learning a potentially productive rule. Experiments on phonotactic pattern learning revealed evidence for feature-based representations via easier learning of single-feature systems. A number of studies show that learning of categories and dependencies occurs only under certain conditions. Many of these studies suggest that infants and children find certain types of linguistic patterns easier to learn than others: patterns that are representationally simpler, that involve local dependencies, and for which multiple cues are available. In this section, we review additional studies that aim to explore the specific ways in which children's learning is constrained by cognition (for reviews of parallel adult studies, see Culbertson 2012; Moreton & Pater 2012a,b).

In the domain of phonology, there is evidence that children are highly sensitive to vowel harmony—a phenomenon characterized by a restriction whereby vowels within a given word must match on some phonological feature, such as vowel height or backness. For example, in Turkish most words are composed of either back vowels or front vowels, not both. Mintz et al. (2018) found that English-acquiring 7-month-olds can pick up on vowel harmony in an artificial language and use it to segment words in continuous speech, even though English does not have this property. One interpretation of these results is that vowel harmony facilitates learning by providing a correlated cue to perceptual grouping: the phonological (or acoustic) similarity between nonadjacent vowels. However, another possibility is that infants have an experience-independent predisposition to attend to vowel harmony. Interestingly, a number of ALL studies have shown a preference for harmony over disharmony (where vowels in a word must be dissimilar) in adults (Pycha et al. 2003; A. Martin & S. Peperkamp, manuscript submitted). This preference correlates with a crosslinguistic generalization: Vowel harmony, not disharmony, is very common across languages.

The cognitive predispositions of learners and their possible link to crosslinguistic typology have been explored using ALL in other domains as well. Hudson Kam & Newport (2009) explored the hypothesis that random, unconditioned variation is rarely found in languages due to a regularization bias on the part of child learners.[4] They taught 5- and 6-year-olds an artificial language in

[4]This is similar to, but distinct from, the idea that children overgeneralize in the following sense: Overgeneralizing a pattern means applying it in some novel context not yet encountered, whereas regularizing a pattern means applying it consistently

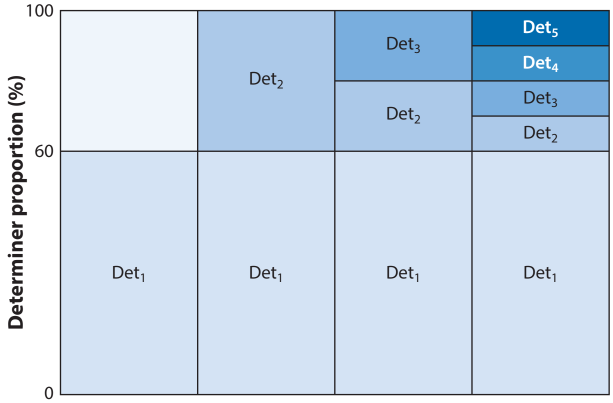


**Figure 4**

Schematic of determiner (Det) variation across conditions in the artificial language learning study by Hudson Kam & Newport (2009). This study investigates how regularization behavior in children and adults is influenced by the level of noise (here unpredictable variation) in determiner usage in the input language.

which nouns were followed by determiners. However, determiner usage in the input was variable. In one condition, the determiner was randomly dropped 40% of the time. In another condition, one determiner was used in 60% of cases but it alternated with two other determiners, each used 20% of the time. Another condition was even noisier, with one main determiner occurring 60% of the time and four alternative determiners occurring 10% of the time each. Importantly, the presence or absence of a particular determiner was not conditioned on anything else (e.g., the particular noun); it was simply random variation. This is shown schematically in **Figure 4**. Children learning under these unsystematic conditions nevertheless produced systematic output patterns: The majority used the main determiner in (almost) all cases or dropped determiners altogether, with the remainder creating some other systematic rule (e.g., consistently adopting one of the alternative determiners).

Interestingly, adult learners were more likely to reproduce the random variation, although they did regularize to some degree. Many follow-up studies have further investigated regularization behavior both in adults and children (e.g., Hudson Kam & Chang 2009, Reali & Griffiths 2009, Perfors & Burns 2010, Smith & Wonnacott 2010). More recently, Schwab et al. (2018) have showed that children learning a novel grammatical gender–like system regularized a probabilistically cued gender marker when they produced utterances in the new language, but not in a task where they were asked to choose which of the two markers to use. This suggests that the regularization bias may be in production rather than in learning or encoding (see Ferdinand et al. 2014 for similar findings in adults). Importantly, Samara et al. (2017) showed that when variation in the input is conditioned—for example, on a sociolinguistic factor such as speaker identity—both adults and children acquire it faithfully. Hudson Kam & Newport (2009) suggested that unconditioned variation was rarely found in natural language because child learners tend to regularize it away. However, conditioned variation is ubiquitous in language; therefore, it is crucial to show that learners do not obliterate all types of variation.

across contexts already encountered or not. For example, a child may know the irregular plural form for some set of words but might overgeneralize the regular plural to a newly encountered noun. However, in this case the child is still using both regular and irregular patterns. By contrast, if the child encountered some set of regular and irregular plurals in the input but systematically used the regular pattern when producing all of them, this would be an example of regularization behavior.

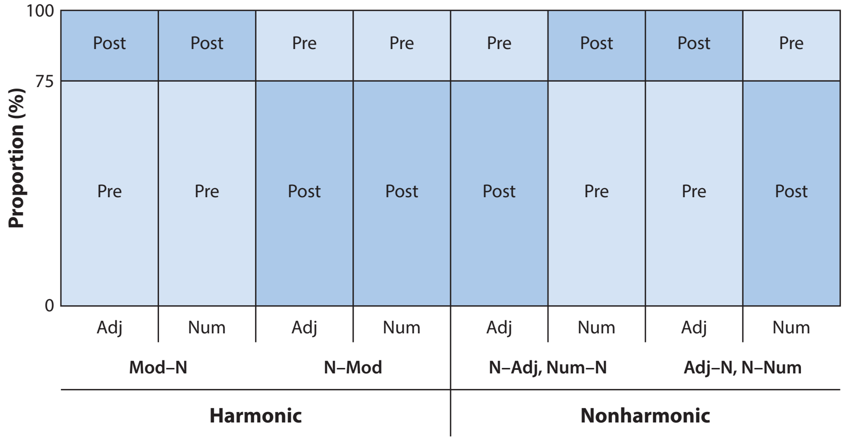


**Figure 5**

Proportion of prenominal and postnominal order for each modifier type in the four conditions used by Culbertson & Newport (2015). This artificial language learning study tests the prediction that adult learners will be more likely to regularize input languages which are predominantly harmonic (both types of modifiers typically on the same side of the noun) than those which are predominantly nonharmonic. Abbreviations: adj, adjective; mod, modifier; num, number word; post, postnominal; pre, prenominal.

Culbertson & Newport (2015) capitalized on children's regularizing behavior to investigate another potential bias in favor of syntactic harmony: a consistent order of heads and dependents. Harmonic languages generally outnumber nonharmonic ones (Greenberg 1963, Dryer 1992, Baker 2001). This potentially reflects a bias for simplicity (Culbertson & Kirby 2016) because harmonic patterns use fewer rules, as do the phonological feature–based generalizations discussed above. For example, languages that place both adjectives and number words consistently before or after the noun they modify (e.g., *red houses* and *two houses* in English) are more common than those that place one before and one after (e.g., *maisons rouges* but *deux maisons* in French). Culbertson & Newport (2015) trained 6- to 7-year-old English-speaking children on an artificial language with variable order of nouns with adjectives and number words. In two harmonic conditions, both types of modifiers either preceded or followed the noun in 75% of the phrases that children were trained on. The remainder of the phrases used the opposite order. In the two nonharmonic conditions, one of the two modifiers preceded the noun and the other followed it in 75% of phrases. This is illustrated in **Figure 5**.

As described by Hudson Kam & Newport (2009), children in this experiment diverged from the input, reducing the overall amount of random variation present in the language. Further, they showed a clear preference in the direction of divergence: Across all conditions, children tended to produce harmonic patterns, even if they were trained on nonharmonic input. Adults showed a similar but weaker harmony bias under these conditions (Culbertson et al. 2012). Braquet & Culbertson (2017) showed that French-speaking children, whose native language is nonharmonic,[5] were nevertheless more likely to regularize a variable harmonic pattern—with postnominal adjectives and numerals—and shifted variable nonharmonic patterns dramatically to conform to that preference. This suggests the intriguing possibility that even experience with a language that is nonharmonic does not overcome a bias for simpler harmonic patterns.

[5]The default adjective order in French is postnominal; however, there is a lexically determined set of adjectives that systematically occur in prenominal position.

Whereas the studies described above rely on random variation in the input to bring out learners' biases, Culbertson & Newport (2017) showed that children innovate harmonic patterns even when trained on deterministically nonharmonic input. Harmonizing errors, in which the learned order of one modifier is generalized to the other, were strongest in 4- to 5-year-olds, weaker in 6- to 7-year-olds, and absent in adults. This suggests that development changes the extent to which learners are able to easily acquire distinct ordering rules across different phrase types. Interestingly, research on regularization points to a similar developmental trajectory: Regularization behavior (or more precisely, systematizing of random variation) drops off around 6–7 years of age (Austin et al. 2006).

These studies provide experimental evidence that cognitive biases (e.g., for simplicity and regularity) shape learning and may come to influence what kinds of linguistic patterns arise and are more faithfully transmitted from generation to generation. Using ALL experiments to explore this phenomenon is a relatively new area of inquiry, but a handful of other studies are worth mentioning here. Bruening et al. (2012) showed that children's sensitivity to changes in the form of a word points to a possible explanation for the apparent preference for suffixal morphology found across the world's languages. The authors used a label extension task, which focuses on how novel labels are used rather than on learning per se. Four-year-old children were shown a single picture and given a novel label for it (e.g., *gep*). Subsequently, they were shown two pictures, one the same as the previous one and one different, and asked to choose which picture a second novel label referred to. This second label was sometimes completely different from the first (e.g., *manse*) and sometimes a modified version of the original label, changed by the addition of either a prefix or a suffix (e.g., *<u>ko</u>gep* or *gep<u>ko</u>*). When the label was completely different, the children rarely chose the same picture: They assumed this label referred to a different object. By contrast, children did extend modified new labels to the same object, but more often when it was modified by the addition of a suffix. This is in line with similar studies with adults, which suggest that words that differ at the end are perceived as more similar than words that differ at the beginning (Hupp et al. 2009).[6]

Several recent studies have built on the classic work by Karmiloff-Smith (1981), who investigated how French children assign grammatical gender to novel nouns. The author taught 3- to 10-year-olds novel nouns denoting female and male alien characters, keeping their grammatical gender ambiguous (e.g., *Voici deux bicrons* 'Here are two *bicrons*,' where *deux* does not carry gender agreement). In some cases, a novel noun with a typically masculine ending (e.g., *-on*) was paired with a female alien, whereas in other cases a novel noun with a typically feminine ending (e.g., *-elle*) was paired with a male alien. The intuitions of adult French speakers indicate that natural gender should override phonology, but children used the definite article that matched the phonological cue. This overreliance on phonology has since been documented in many languages (Mills 1985, Perez-Pereira 1991, Demuth 2000, Demuth & Ellis 2008, Mariscal 2009, Rodina & Westergaard 2012). Gagliardi & Lidz (2014) replicate this in the Tsez (Nakh-Daghestanian) language, where the semantic cues are demonstrably more reliable predictors of noun category than the phonological cues are (including in child-directed speech). This suggests that learners may be biased to attend to internal cues such as noun phonology rather than external cues such as noun meaning.

As discussed in Section 3, infants and young children appear to construct linguistic representations gradually, learning about phonological features of their language very early on, and later grouping sounds into words and words into more complex phrases. Culbertson et al. (2017,

---

[6]This is related to the more specific claim that having lexical information at the beginning allows words to be uniquely identified earlier (Hawkins & Cutler 1988).

2018) investigated whether young children's curious overreliance on phonology might reflect this kind of staged acquisition. They argue that children may begin forming representations of noun categories purely on the basis of phonological dependencies (e.g., between the form of a determiner and a noun) before they have learned noun meanings. If early representations are based on phonology, and semantic information is integrated later, then children may go through a period of apparent overreliance on phonology without having a phonological bias per se. Culbertson et al. (2017) showed that adult learners of an artificial gender system indeed rely more on phonology if they have access to phonological information first and semantics later. Culbertson et al. (2018) replicated this effect with children (using a paradigm very similar to the one used by Schuler et al. 2016). However, they also found evidence that children rely more on phonology than adults even when both sources of information are simultaneously present. These results, again, support the idea that early learning is shaped both by learners' biases and by features of the input.[7]

Finally, in the domain of semantics, Hunter & Lidz (2012) showed that children's acquisition of novel determiners (or quantifiers) mirrors a typological restriction on determiner semantics. Most, if not all, natural language determiners exhibit a property called conservatism (Keenan & Stavi 1986). Intuitively, words like *some*, *all*, *every*, and *most* define relations between sets: *Some dogs are brown* is a relation between the set of dogs and the set of brown things. Conservatism states that in evaluating this relation between sets, it suffices to only consider dogs. In other words, in determining whether the statement *some dogs are brown* is true, one need consider only the set of dogs, not the set of all things that are brown. In principle, one could easily conceive of a determiner whose meaning is not conservative. Take the fictional determiner *equi*, where a sentence such as *equi dogs are brown* is true if and only if the set of dogs is equal to the set of brown things. To evaluate this, one must consider both the size of the set of dogs and the size of the set of brown things. However, these types of meanings are not found as grammaticalized determiners (the concepts can be expressed, but not in a single word).

To test whether this restriction on determiner semantics reflects a cognitive bias on the part of learners, Hunter & Lidz (2012) taught 5-year-olds one of two novel determiners. The first encoded the conservative meaning 'not all' (e.g., *gleeb girls are on the beach* was only true if not all girls were on the beach). The second encoded the nonconservative meaning that is the mirror image of 'not all' (e.g., *gleeb girls are on the beach* was only true if not all the beachgoers were girls). The latter is nonconservative because it requires considering the set of beachgoers. Training was implemented as a card sorting task: The experimenter had a set of cards and sorted some of them according to whether *gleeb girls are on the beach*. The child was then asked to help the experimenter sort the remaining cards. Children sorted more cards correctly in the conservative determiner condition. This suggests that children find nonconservative determiners more difficult to learn, providing a window into what semantic meanings learners find more complex and a possible explanation for why such words are not found in languages.

## 6. NEXT STEPS AND NEW DIRECTIONS

In the preceding sections, we have reviewed highlights from research using ALL to better understand the nature of children's early representations and what shapes them. In our view, a critical next step is to investigate how the representations themselves and what constrains them

[7]For related research on how to improve adult second-language acquisition of gender by making adults learn more like children, see Arnon & Ramscar (2012) and Siegelman & Arnon (2015). The basic idea is that children's successful learning of noun gender is facilitated by the fact that they do not reliably segment markers from the noun during early learning.

change across development. How does the likelihood of generalization from a particular input distribution change as a function of cognitive and linguistic development? Do the cognitive biases we see in young children persist into adulthood, and how does experience with a particular language affect this? These questions require comparing across a wider range of age groups—infants, young children, adolescents, and adults. Further, the vast majority of ALL studies have focused on English-acquiring infants and children. The phenomena under investigation are not specific to English, and focusing on this population prevents us from fully understanding how development is shaped by linguistic experience and will undoubtedly lead us to miss out on critical issues in crosslinguistic acquisition.

Relatedly, we have discussed both properties of the input and properties of the learner that affect what is learned. In general, however, very little work uses ALL to explore the intake (i.e., the information available in an input that a child is able to use at any given point in the acquisition process). Just as representations change, the intake changes as learners' knowledge becomes more sophisticated. This applies across domains but has been central to work on acquisition of noun categories (e.g., Gagliardi & Lidz 2014; Culbertson et al. 2016, 2018). If early acquisition is based on phonological information and only later incorporates semantic information, this not only will affect what early representations look like but also may continue to shape later learning. Along the same lines, how is knowledge of long-distance dependencies built up? If the intake initially involves less abstract, more local information, then more abstract knowledge of dependencies may be affected by how those early representations are acquired (e.g., Culbertson et al. 2016).

These next steps depend on a tight connection between natural language acquisition and ALL results. Observations and hypotheses from these domains should feed each other. Most obviously, the fact that a learner is able to acquire a particular pattern in a stripped-down ALL setting does not necessarily mean that learners can do this in the wild. There may be features of a richer linguistic or social environment that block or promote learning, and this is part of what we want to know as language acquisition researchers. Similarly, we have seen here that some patterns are more difficult to learn in the lab than others, resulting in errors of a particular type—for example, promoting word-order harmony. However, in order to connect these findings to language typology, we need to better understand the natural language linking mechanisms. Under what circumstances are learners' errors propagated over time in a way that would affect the linguistic system of the community?

## 7. CONCLUSION

We have reviewed key research using ALL to investigate the nature of early linguistic representations and how they are influenced by both the structure of the input and the cognitive features of the learner. These studies have revealed evidence for not only early abstraction and generalization in both phonology and syntax but also conditions or constraints on these processes. Similarly, ALL studies have shed light on the conditions under which children develop knowledge about grammatical categories, constituents, and dependencies, at the same time revealing that some types of representations are easier to acquire than others (e.g., simpler rules, adjacent dependencies). However, particular features of the input were shown to interact with this; for example, correlated cues make otherwise-difficult categories and dependencies easier to learn. These studies, along with recent work on harmony, suffixation, and other apparent crosslinguistic regularities, illustrate how ALL can be used to understand language acquisition and the potential link between cognition and language typology. Whereas these methods, in conjunction with work on natural language, have pushed the discipline forward significantly in the past several decades, there are a number of interesting directions which we have argued future work should take, including exploring the impact of linguistic and cognitive development on abstraction, generalization, and cognitive biases.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.



## LITERATURE CITED

Adger D. 2003. *Core Syntax*. Oxford, UK: Oxford Univ. Press

Arnon I, Ramscar M. 2012. Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition* 122:292–305

Aslin RN. 2007. What's in a look? *Dev. Sci.* 10:48–53

Aslin RN, Newport EL. 2012. Statistical learning from acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci.* 21:170–76

Aslin RN, Saffran JR, Newport EL. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9:321–24

Austin AC, Newport EL, Wonnacott E. 2006. *Predictable versus unpredictable variation: regularization in adult and child learners*. Paper presented at the 31st Annual Boston University Conference on Language Development, Boston, MA

Baker CL. 1979. Syntactic theory and the projection problem. *Linguist. Inq.* 10:533–81

Baker M. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books

Berko J. 1958. *The child's learning of English morphology*. PhD thesis, Radcliffe Coll., Cambridge, MA

Braine MDS. 1963. On learning the grammatical order of words. *Psychol. Rev.* 70:323

Braine MDS. 1987. What is learned in acquiring word classes: a step toward an acquisition theory. In *Mechanisms of Language Acquisition*, ed. B MacWhinney, pp. 65–87. Hillsdale, NJ: Erlbaum

Braine MDS, Brody RE, Brooks PJ, Sudhalter V, Ross JA, et al. 1990. Exploring language acquisition in children with a miniature artificial language: effects of item and pattern frequency, arbitrary subclasses, and correction. *J. Mem. Lang.* 29:591–610

Braquet G, Culbertson J. 2017. Harmony in a non-harmonic language: word order learning in French children. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, ed. G Gunzelmann, A Howes, T Tenbrink, E Davelaar, pp. 156–61. Austin, TX: Cogn. Sci. Soc.

Brooks PJ, Braine MDS, Catalano L, Brody RE, Sudhalter V. 1993. Acquisition of gender-like noun subclasses in an artificial language: the contribution of phonological markers to learning. *J. Mem. Lang.* 32:76–95

Bruening P, Brooks P, Alfieri L, Kempe V, Dabasinskiene I. 2012. Children's tolerance of word-form variation. *Child Dev. Res.* 2012:401680

Chomsky N, Halle M. 1968. *The Sound Pattern of English*. New York: Harper & Row

Crain S, Nakayama M. 1987. Structure dependence in grammar formation. *Language* 63:522–43

Cristià A, Seidl A. 2008. Is infants' learning of sound patterns constrained by phonological features? *Lang. Learn. Dev.* 4:203–27

Culbertson J. 2012. Typological universals as reflections of biased learning: evidence from artificial language learning. *Lang. Linguist. Compass* 6:310–29

Culbertson J, Gagliardi A, Smith K. 2017. Competition between phonology and semantics in noun class learning. *J. Mem. Lang.* 92:343–58

Culbertson J, Jarvinen H, Haggarty F, Smith K. 2018. Do children privilege phonological cues in noun class learning? In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, ed. C Kalish, M Rau, R Zhu, T Rogers. Austin, TX: Cogn. Sci. Soc. In press

Culbertson J, Kirby S. 2016. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Front. Psychol.* 6:1964

Culbertson J, Koulaguina E, Gonzalez-Gomez N, Legendre G, Nazzi T. 2016. Developing knowledge of non-adjacent dependencies. *Dev. Psychol.* 52:2174–83

Culbertson J, Newport EL. 2015. Harmonic biases in child learners: in support of language universals. *Cognition* 139:71–82

Culbertson J, Newport EL. 2017. Innovation of word order harmony across development. *Open Mind Discov. Cogn. Sci.* 1:91–100

Culbertson J, Smolensky P, Legendre G. 2012. Learning biases predict a word order universal. *Cognition* 122:306–29

Demuth K. 2000. Bantu noun class systems: loanword and acquisition evidence of semantic productivity. In *Systems of Nominal Classification*, ed. G Senft, pp. 270–92. Cambridge, UK: Cambridge Univ. Press

Demuth K, Ellis D. 2008. Revisiting the acquisition of Sesotho noun class prefixes. In *Crosslinguistic Approaches to the Psychology of Language: Festschrift for Dan Slobin*, ed. J Guo, E Lieven, S Ervin-Tripp, N Budwig, K Nakamura, pp. 93–104. Mahwah, NJ: Erlbaum

de Vries MH, Monaghan P, Knecht S, Zwitserlood P. 2008. Syntactic structure and artificial grammar learning: the learnability of embedded hierarchical structures. *Cognition* 107:763–74

Dryer M. 1992. The Greenbergian word order correlations. *Language* 68:81–183

Erickson LC, Thiessen ED. 2015. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.* 37:66–108

Ettlinger M, Morgan-Short K, Faretta-Stutenberg M, Wong P. 2016. The relationship between artificial and second language learning. *Cogn. Sci.* 40:822–47

Ferdinand V, Kirby S, Smith K. 2014. Regularization in language evolution: on the joint contribution of domain-specific biases and domain-general frequency learning. In *Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)*, ed. EA Cartmill, S Roberts, H Lyn, H Cornish, pp. 435–36. Singapore: World Sci.

Friederici AD, Mueller JL, Oberecker R. 2011. Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLOS ONE* 6:e17920

Gagliardi A, Lidz J. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90:58–89

Gerken L. 2004. Nine-month-olds extract structural principles required for natural language. *Cognition* 93:B89–96

Gerken L. 2006. Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition* 98:B67–74

Gerken L, Wilson R, Lewis W. 2005. Infants can use distributional cues to form syntactic categories. *J. Child Lang.* 32:249–68

Gómez RL. 2002. Variability and detection of invariant structure. *Psychol. Sci.* 13:431–36

Gómez RL, Gerken L. 1999. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70:109–35

Gómez RL, Gerken L. 2000. Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4:178–86

Gómez RL, Lakusta L. 2004. A first step in form-based category abstraction by 12-month-old infants. *Dev. Sci.* 7:567–80

Gómez RL, Maye J. 2005. The developmental trajectory of nonadjacent dependency learning. *Infancy* 7:183–206

Greenberg J. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, ed. J Greenberg, pp. 73–113. Cambridge, MA: MIT Press

Hawkins JA, Cutler A. 1988. Psycholinguistic factors in morphological asymmetry. In *Explaining Language Universals*, ed. JA Hawkins, pp. 280–317. Oxford, UK: Blackwell

Hawthorne K, Rudat L, Gerken L. 2016. Prosody and the acquisition of hierarchical structure in toddlers and adults. *Infancy* 21:603–24

Hayes B, Kirchner R, Steriade D. 2004. *Phonetically Based Phonology*. Cambridge, UK: Cambridge Univ. Press

Hohle B, Schmitz M, Santelmann LM, Weissenborn J. 2006. The recognition of discontinuous verbal dependencies by German 19-month-olds: evidence for lexical and structural influences on children's early processing capacities. *Lang. Learn. Dev.* 2:277–300

Houston-Price C, Nakai S. 2004. Distinguishing novelty and familiarity effects in infant preference procedures. *Infant Child Dev*. 13:341–48

Hudson Kam C, Chang A. 2009. Investigating the cause of language regularization in adults: memory constraints or learning effects? *J. Exp. Psychol. Learn. Mem. Cogn.* 35:815

Hudson Kam C, Newport E. 2009. Getting it right by getting it wrong: when learners change languages. *Cogn. Psychol.* 59:30–66

Hunter T, Lidz J. 2012. Conservativity and learnability of determiners. *J. Semant.* 30:315–34
Hupp JM, Sloutsky VM, Culicover PW. 2009. Evidence for a domain-general mechanism underlying the suffixation preference in language. *Lang. Cogn. Process.* 24:876–909
Johnson EK, Jusczyk PW. 2001. Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44:548–67
Johnson EK, Tyler MD. 2010. Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13:339–45
Jusczyk PW, Cutler A, Redanz NJ. 1993. Infants' preference for the predominant stress patterns of English words. *Child Dev*. 64:675–87
Jusczyk PW, Luce P, Charles-Luce J. 1994. Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* 33:630–45
Karmiloff-Smith A. 1981. *A Functional Approach to Child Language: A Study of Determiners and Reference*. New York: Cambridge Univ. Press
Keenan EL, Stavi J. 1986. A semantic characterization of natural language determiners. *Linguist. Philos.* 9:253–326
Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A, Gerken L. 1995. The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev.* 18:111–16
Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606–8
Langus A, Marchetto E, Bion RAH, Nespor M. 2012. Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66:285–306
Lany J, Gómez RL. 2008. Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol. Sci.* 19:1247–52
Lightfoot D. 1999. *The Development of Language: Acquisition, Change, and Evolution*. New York: Blackwell
Marcus GF, Vijayan S, Rao SB, Vishton PM. 1999. Rule learning by seven-month-old infants. *Science* 283:77–80
Mariscal S. 2009. Early acquisition of gender agreement in the Spanish noun phrase: starting small. *J. Child Lang.* 36:143–71
Mills AE. 1985. The acquisition of German. In *The Crosslinguistic Study of Language Acquisition, vol. 1: The Data*, ed. DI Slobin, pp. 141–254. Hillsdale, NJ: Erlbaum
Mintz TH. 1996. *The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories*. PhD thesis, Univ. Rochester, Rochester, NY
Mintz TH, Walker RL, Welday A, Kidd C. 2018. Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition* 171:95–107
Misyak JB, Christiansen MH. 2012. Statistical learning and language: an individual differences study. *Lang. Learn.* 62:302–31
Moreton E, Pater J. 2012a. Structure and substance in artificial-phonology learning, part I: structure. *Lang. Linguist. Compass* 6:686–701
Moreton E, Pater J. 2012b. Structure and substance in artificial-phonology learning, part II: substance. *Lang. Linguist. Compass* 6:702–18
Morgan JL, Meier RP, Newport EL. 1987. Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cogn. Psychol.* 19:498–550
Morgan JL, Meier RP, Newport EL. 1989. Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *J. Mem. Lang.* 28:360–74
Morgan JL, Newport EL. 1981. The role of constituent structure in the induction of an artificial language. *J. Verbal Learn. Verbal Behav.* 20:67–85
Morgan JL, Saffran JR. 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Dev*. 66:911–36
Newport EL. 1990. Maturational constraints on language learning. *Cogn. Sci.* 14:11–28
Newport EL, Aslin RN. 2004. Learning at a distance I: statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48:127–62

Onnis L, Monaghan P, Richmond K, Chater N. 2005. Phonology impacts segmentation in online speech processing. *J. Mem. Lang.* 53:225–37

Peperkamp S, Le Calvez R, Nadal J-P, Dupoux E. 2006. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101:B31–41

Perez-Pereira M. 1991. The acquisition of gender: what Spanish children tell us. *J. Child Lang.* 18:571–90

Perfors A, Burns N. 2010. Adult language learners under cognitive load do not over-regularize like children. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, ed. S Ohlsson, R Catrambone, pp. 2524–29. Austin, TX: Cogn. Sci. Soc.

Petersson KM, Forkstam C, Ingvar M. 2004. Artificial syntactic violations activate Broca's region. *Cogn. Sci.* 28:383–407

Pullum G, Scholz BC. 2002. Empirical assessment of stimulus poverty arguments. *Linguist. Rev.* 19:9–50

Pycha A, Nowak P, Shin E, Shosted R. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, ed. M Tsujimura, G Garding, pp. 101–14. Somerville, MA: Cascadilla

Reali F, Griffiths TL. 2009. The evolution of frequency distributions: relating regularization to inductive biases through iterated learning. *Cognition* 111:317–28

Reber AS. 1967. Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6:855–63

Reber AS. 1969. Transfer of syntactic structure in synthetic languages. *J. Exp. Psychol.* 81:115–19

Reber AS. 1989. Implicit learning and tacit knowledge. *J. Exp. Psychol. Gen.* 118:219–35

Reeder PA, Newport EL, Aslin RN. 2013. From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cogn. Psychol.* 66:30–54

Rodina Y, Westergaard M. 2012. A cue-based approach to the acquisition of grammatical gender in Russian. *J. Child Lang.* 39:1077–106

Romberg AR, Saffran JR. 2010. Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1:906–14

Saffran JR. 2003. Statistical language learning mechanisms and constraints. *Curr. Dir. Psychol. Sci.* 12:110–14

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–28

Saffran JR, Thiessen ED. 2003. Pattern induction by infant language learners. *Dev. Psychol.* 39:484

Samara A, Smith K, Brown H, Wonnacott E. 2017. Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cogn. Psychol.* 94:85–114

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

Schuler K, Yang C, Newport E. 2016. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, ed. A Papafragou, D Grodner, D Mirman, J Trueswell, pp. 2321–26. Austin, TX: Cogn. Sci. Soc.

Schwab JF, Lew-Williams C, Goldberg AE. 2018. When regularization gets it wrong: Children over-simplify language input only in production. *J. Child Lang.* 45:1054–72

Siegelman N, Arnon I. 2015. The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *J. Mem. Lang.* 85:60–75

Smith K, Wonnacott E. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116:444–49

Tomasello M. 2000. The item-based nature of children's early syntactic development. *Trends Cogn. Sci.* 4:156–63

Werker JF, Tees RC. 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7:49–63

Wonnacott E. 2011. Balancing generalization and lexical conservatism: an artificial language study with child learners. *J. Mem. Lang.* 65:1–14

Wonnacott E. 2013. Learning: statistical mechanisms in language acquisition. In *The Language Phenomenon*, ed. PM Binder, K Smith, pp. 65–92. New York: Springer

Wonnacott E, Brown H, Nation K. 2017. Skewing the evidence: the effect of input structure on child and adult learning of lexically based patterns in an artificial language. *J. Mem. Lang*. 95:36–48
Wonnacott E, Newport EL, Tanenhaus MK. 2008. Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cogn. Psychol.* 56:165–209
Yang CD. 2004. Universal grammar, statistics or both? *Trends Cogn. Sci.* 8:451–56
Yang CD. 2005. On productivity. *Linguist. Var. Yearb.* 5:265–302