Infant artificial language learning and language acquisition

Rebecca L. Gómez and LouAnn Gerken

Language acquisition is one of the most complex learning tasks imaginable. The daunting nature of the undertaking arises from conflicting pressures to generalize beyond the stimuli encountered without generalizing too far. For example, it has been observed that children never erroneously transform a statement like ‘The man who is tall is Sam’ into the question ‘Is the man who tall is Sam?’, by moving the subordinate clause verb rather than the main verb to the front of the sentence. The lack of such errors has been taken as evidence that children never consider rules based solely on linear order in sentences, such as ‘move the first verb to the front of the sentence’. The computational and logical difficulties raised by these conflicting pressures have caused many researchers to conclude that language is not learnable by an unspecialized learning device and that children must be born with some number of built-in constraints for deciding when and how to generalize from the stimuli they encounter. This view of a constrained language learner has dominated the field for the past 30 years or so. However, recent advances in cognitive science are causing us to reconsider the type and degree of constraints placed on the learner. Of particular interest, and the focus of this article, are recent studies on infants’ ability to acquire information about miniature artificial languages after very brief exposure.

The complexity of natural language makes it exceedingly difficult to isolate factors responsible for language learning. In English, words like the and a tend to be accompanied by intonational patterns involving brief pausing and reduced stress. There has been considerable speculation that such cues might help learners discover the syntax of their native language, but whether these features of sentences and clauses are sensitive to these features of sentences and clauses, we do not know. Thus, the focus of this article is on infants’ ability to acquire information about miniature artificial languages after very brief exposure. Language researchers have thus turned to artificial languages as a means of obtaining better control over the input to which learners are exposed. Artificial languages can be designed to...
Box 1. Learning in utero

Artificial-language studies with infants demonstrate the presence of remarkably sophisticated learning abilities by seven, eight and 12 months of age (Refs a–d). Such findings instigate many new studies have led to exciting discoveries regarding the actual mechanisms involved. Training infants on certain features of their training passage in utero is likely to result from a gradual accumulation of knowledge at the presence of remarkably sophisticated learning abilities by seven, eight and 12 months of age (Refs a–d). Such findings instigate new studies have led to exciting discoveries regarding the actual mechanisms involved. Training infants on certain features of their training passage in utero is likely to result from a gradual accumulation of knowledge at the.

References

Box 2. Nativist and empiricist views of language acquisition

Although the position of the present review takes as fundamental the assumption that human infants are born with a language-specific learning device (Ref. 4), according to this view, the structure of language learning, certain aspects of the formal structure of language (thought to be absolute and universal) are genetically specified so that acquiring one’s target language is an immediate logical dilemma as framed, increasing evidence suggests that the empirical problem faced by children is not so intractable.

It is important to point out that although these views differ in their approach to the language problem, each side acknowledges the contribution of the other. One of the assumptions of the nativist view is that once a parameter is set for a particular rule, that rule will be applied across a wide range of category instances. However, fine-grained examination of children’s utterances suggests that instead of generalizing widely, children first use a small number of structures they use tend to be the ones most frequent in their input. Thus, the ability to acquire more specific data should aid in circumventing limitations in both approaches.

Central to this view is the argument from the poverty of the stimulus, that linguistic input is too impoverished and learning mechanisms too weak to otherwise explain how young children acquire language (where convergence on a universal grammar is thought to underly linguistic productivity). Convergence on the form of linear order, such as ‘nouns first to the front of the sentence’, because they never erronously transform sentences like ‘The man who is tall is Sam’ into ungrammatical counterparts like ‘Is the man who is tall Sam?’

Given that children hear many simple instances that might lead them to form a rule based on linear order (e.g., ‘the boy is tall’, ‘the girl is tall’), how do we explain the lack of errors in sentences with a subclasses clause (and hence two verbs)? Given the conflicting evidence available in the environment, a classic answer is to assume children are innately constrained to consider the hierarchical organization (or structural dependencies) of syntactic phrases as opposed to linear word ordering. Arguments such as these gain considerable momentum from Gold’s proof showing that certain classes of language are not learnable without some kind of constraint on the hypotheses learners are willing to entertain (Ref. 17).

An alternative, empiricist view sees the learner as a blank slate, equipped with general associative learning mechanisms (Ref. 18). According to this view, learning might be constrained by human information-processing abilities, but is not limited to the specific domain of language. Although the nativist view has dominated for many years, recent advances in cognitive science suggest that the assumptions underlying this view might have been overly restrictive (Ref. 18). First, far from being impoverished, the language children hear is rich in structural regularities (Ref. 19). Such regularities aid learning in humans (Ref. 20) and in neural networks. Second, neural networks have far outstripped early conceptions (Ref. 10) of associative learning, especially with respect to their ability of capturing key aspects of linguistic behavior (Ref. 10).

Where these models differ from nativist proposals is in their emphasis on learning as a stochastic process over distributed input rather than one involving manipulation of discrete symbols. For example, Rohde and Plant have recently demonstrated how such architectures learn without explicit negative feedback (Ref. 10). The fact that human infants also acquire syntactical regularities (Ref. 22) suggests a degree of overlap in at least some of the mechanisms involved. Researchers have also begun to argue that although Gold’s proof applies under the assumption that all learners converge on one true target, it does not apply under the assumption that the target is mechanistically defined (Ref. 10).

Thus, consistent cue is that syllables within words usually have higher transitional probabilities than syllables spanning words (a ‘transitional probability’, the conditional probability of Y given X, is calculated by normalizing the co-occurrence frequency of X and Y by the frequency of X) (Ref. 12).

For example, in the listener’s experience, the likelihood that pretty will follow in the phrase pretty baby is much higher than the likelihood that ke will follow ty. Why? Many words other than baby can follow pretty (e.g. pretty daggis, pretty monkey, pretty girl, pretty flower or pretty dolly).
Saffran, Ades, and Newport investigated whether infants could use statistical regularities to identify words in running speech. In their study, eight-month-old infants listened to two minutes of continuous speech consisting of four tri-syllabic nonsense words strung together in random order (e.g. bidakupadotigolabubidakutupi). Infants were then tested to determine whether they would discriminate two of the familiarized words (e.g. tripe and gelale) from two nonwords (alpafa and inale). Infants’ listening preferences for different stimuli were measured using the head-turn preference procedure. Stimuli in this procedure are presented auditorily from the infant’s left or right side. The amount of time the infant orients toward the source of sound is taken as the dependent measure. Words and non-words were chosen from the same syllable set, but differed in terms of the transitional probabilities of l and non-words having mean transitional probabilities of 0 and non-words having mean transitional probabilities of 0. The only cue to whether or not a stimulus was a word was the difference in mean transitional probabilities, and so discrimination would demonstrate sensitivity to such probabilities. Infants, in fact, showed differential attention to familiar and unfamiliar syllable combinations, suggesting the presence of a fairly sophisticated statistical learning mechanism. Later studies demonstrated that infants were also sensitive to transitional probabilities over tone sequences, suggesting that this learning mechanism was more general than one dedicated solely to processing linguistic stimuli. Whether infants will go on to treat constituents extracted from speech as lexical items is still open, but it is certainly a question that can be investigated empirically.

**Words in sequence**

In addition to segmenting words in running speech, listeners must also acquire the legal ordering of words in sentences. To determine whether infants could learn ‘grammatical’ word order, Gómez and Gerken exposed 12-month-olds to a subset of strings produced by one of two grammars (see Fig. 1). Note that although word order is constrained by these grammars, there is still considerable variability in terms of the endings of words in sentences. For example, in Grammar 1, PEL can occur in first position (PEL-TAM-RUD), second position (VOT-PEL-JIC-RUD-TAM), second and third position (VOT-PEL-JIC-RUD-TAM), or not at all (e.g. VOT-JIC-RUD-TAM). Similarly, JIC occurs after either VOT, PEL, or TAM, but its position varies as a function of whether the sentence begins with PEL or VOT, whether PEL occurs after VOT or after TAM, and whether PEL repeats in the string. Also, brief exposure to a subset of strings of their training grammar (between 50 and 127 seconds), infants were given a short play break, and then were tested to see if they would discriminate new strings from the two grammars. Importantly, both grammars began and ended with the same words and contained the same vocabulary. They differed, however, in terms of the ordering of word pairs. For instance, the transition T Amph/JIC found in Grammar 2 never occurred in Grammar 2. Likewise, VOT-RUD found in Grammar 2 never occurred in Grammar 1. Infants listened longer to new strings from their training grammar than to strings from the other grammar, regardless of which grammar 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 encountered during training. Although demonstrating that learning was not confined to memory for particular strings, but rather generalized to novel strings with familiar co-occurrence patterns. This learning task all the
more remarkable given that it occurred after less than two minutes exposure and was retained over a short delay.

It is likely that the statistical learning mechanism documented by Saffran and colleagues12 also explains the learning in these studies. Importantly however, learning is not so static as to prohibit recognition of grammatical word combinations in novel sentences.

Words in abstract patterns

Although sensitivity to word order is necessary for tracking sequential information in sentences, learners must ultimately abstract beyond the ordering of specific words. It is with this aim that researchers have begun investigating early abstraction abilities. For instance, Gómez and Gerken11 exposed infants to a subset of strings produced by one of the two grammars shown in Fig. 1. Instead of using the vocabulary depicted in the figure, the training set consisted of JED, FIM, TUP, DAK and SOG. The test strings, however, were constructed using the vocabulary VOT, PEL, JIC, RUD and TAM. To give an example, infants trained on Grammar 1 heard strings like FIM-SOG-FIM-FIM-TUP and were tested on new strings like VOT-PEL-PEL-JIC. Thus, although constraints on grammatical word ordering remained constant, vocabulary did not. Critically, because test strings were instantiated in new vocabulary, learners could not distinguish the two grammars based on transitional probabilities between remembered word pairs. This task was all the more difficult because the subset of strings used during training did not overlap with the subset of grammatical strings used at test. That is, none of the underlying strings occurred in both training and at test. Infants discriminated grammatical from ungrammatical strings despite the change in vocabulary and despite the fact that none of the underlying test strings were encountered during training, suggesting that they had abstracted some aspect of grammatical structure above and beyond pairs of specific elements. This ability does not appear to be domain specific, at least with respect to adult learners. Adults trained on visually presented consonant and symbol strings generalized to auditorily presented tone and CVC sequences (and vice versa)17,18. It remains to be seen whether such learning will prove to be domain general for younger learners.

In a similar series of studies, Marcus et al.19 exposed seven-month-olds to three minute speech samples of strings with ABA (a-d-e) and ABB (a-d-d) word patterns. In these studies the underlying pattern was the same for training and for test, however, the vocabulary was different. Infants were subsequently able to discriminate strings with the training pattern from those with a different pattern (e.g. ba-po-ba versus ba-po-po), despite the change in vocabulary. These results were important for demonstrating that younger infants can also abstract beyond specific word order. Marcus et al. further interpreted these findings as evidence that infants are acquiring algebra-like rules (involving substitution of arbitrary elements in abstract variables), an example from language would be the substitution of any plural noun phrase for 'the three daxels strolled through the park'14. Marcus has argued that systems sensitive only to statistical regularities (namely connectionist architectures) are, in principle, incapable of such abstraction14–15. Arguments against this interpretation (as well as several demonstrations in favor of a statistical learning account of such abstraction) have been mounted by a number of researchers. Thus, although the issue of whether infants are abstracting by means of rules or statistical regularities is still open to debate17,18, there is no doubt that infants can generalize beyond specific word order. Having demonstrated such abstraction, we must next ask how central it is to acquiring the syntax of one’s native language.

Limitations of pattern-based representations

The infant abstraction abilities documented thus far have in common that grammatical and ungrammatical strings were distinguishable by differences in patterns of identical elements (e.g. ABB, ABA, ABCA and ABAAC)11,14. No doubt identity is salient for learners. When absent, infants and adults no longer generalize, providing support for the hypothesis that identity underlies this abstraction5–6. Gómez et al.11 exposed learners to a grammar containing strings with one repeating element versus a grammar with no identical elements. Learners acquired robust knowledge of sequential dependencies (as
reflected in their ability to discriminate grammatical from ungrammatical strings in their training vocabulary). However, such knowledge did not factor into their ability to generalize to new vocabulary. Abstraction beyond specific word order only occurred for learners trained on the grammar with re- peating elements. Such abstraction could be limited, how- ever, with respect to acquiring syntax. The key to under- standing this point lies in a contrast between what we will call pattern-based and category-based abstraction.

Pattern-based abstraction can be described in terms of relational operations (e.g., identity, greater-than or less-than) over physical stimuli in sequence. For example, recognizing ba-po-yi and ba-po-yi as instances of the pattern ABA entails noting that the first and last syllables in sequence are physi- cally identical. It is perhaps easier to understand this distinc- tion in the context of classic studies from the animal literature. For example, chimpanzees, rats and chicks are able to evaluate relational patterns in training stimuli (e.g., luminance, > luminance, ) and generalize to untrained stimuli. Furthermore, starlings trained to respond to ascend- ing tone sequences generalize to new ascending sequences created by various transformations of the training stimuli (where an ascending sequence can be described by a series of relations in which the pitch of sequence element is greater than element ). In each of these examples, a re- lation is abstracted by comparing the perceptual character- istics of each element in the physical array to those of the other elements. In this way, such relations are perceptually bound.

Category-based generalization, by contrast, involves op- erations over abstract rather than perceptually bound vari- ables. Compare the pattern-based representation ABA with the category-based representation Noun-Verb-Noun. Although superficially similar, these examples differ along a critical dimension. Recognizing ABA and Noun-Verb-Noun both involve identity, but in the former case, the relation is perceptually bound, whereas in the latter the identity relation holds over abstract categories and thus is at least one step re- moved from physical identity. That is, abstracting the pat- tern ABA from ba-po-yi involves noting that the first and third elements in a sequence are physically identical. With category-based generalization, however, learners must iden- tify the first and third elements as members of the abstract category ‘noun’. These determinations cannot be based on a subset of the pairings. For instance, learners might hear M, paired with N, and N, M, paired with N, and M, paired with N. The question is whether they will generalize to the untrained pairings (left marked with ). If they have learned the dependencies between categories, then they should accept new grammatical pairings such as Mn, while rejecting ungrammatical ones (e.g., M-Q).

The ability to abstract over categories is fundamental to lin- guistic productivity. A learner who identifies a novel word as belonging to a particular category has immediate access to all of the rules involving that category. Even very young learners are privy to such information. Pre-school children seeing and hearing ‘Here is a wug. Now there are two of them’, and asked to complete the sentence ‘There are two ______’, respond with the answer ‘wugs’.

Category-based abstraction has been of particular inter- est to researchers investigating language learning mechanisms with older learners, and has focused on the problem of how learners acquire relations between grammatical classes. For example, English-speaking children need to learn that the determiners and and a precede nouns and not verbs, whereas auxiliaries like and is not only limited to the present tense, but not nouns. This problem can be conceptualized in terms of filling in the cells of matrices (such as the ones shown in Fig. 2), where learners must acquire the knowledge that MN and PQ are legal sentences of a language, but MQ and PN are not. In these studies, learners are exposed to most, but not all, grammatical pairings during training to see whether they will generalize to new grammatical pairs not seen. For example, fringe and McDonald augmented a subset of the members from the
Box 3. Systematically related cues and learning

Research on category-based abstraction has been critical for demonstrating the importance of systematically related cues in abstraction of language-like categories (Refs a–c). However, it should come as no surprise that learners rely heavily on correlated systematic cues to learn group membership (e.g. noun or verb). Presumably, learners group N- and Q-words into different categories based on their distinguishing features. These features are then used to group M- and P-elements. Once categorized, learners can use the knowledge that M1 pairing with N1 to infer that it also pairs with N2, to infer that it also pairs with N3 (Ref. 40).

The results of these studies are instructive in more than one way. First, they demonstrate that humans are not unconstrained learners. People do not make discrete arbitrary distinctions. Abstraction results only when there is sufficient evidence to distinguish the categories in question. This should not be surprising given previous work on the importance of correlated cues in language learning (Box 3). However, this fact about human learning is also important for circumscribing the nature of the acquisition mechanisms proposed. For example, overly powerful models have assumed that learning involves abstraction of arbitrary structure10,12, when in fact many of the categories found in natural language (such as gender, declension and conjugation classes) are rich in systematic cues to class membership13. For example, in Spanish, feminine nouns often end in -a and masculine nouns in -o. In Hebrew, nouns ending in -a and -e are often feminine.

References


Now that we have some understanding of the requirements for inducing category-based abstraction in adults, the next step will be to begin investigating how younger learners master such abstraction. Other important issues have to do with whether such learning is rule based or associative in nature and whether learners can induce category structure based on a more limited set of examples (e.g., exhibiting characteristics found in child-directed speech). We are currently investigating these issues in our joint laboratories with infants and adults.

Conclusion and implications

We have reviewed a number of studies investigating the learning abilities exhibited by infants and adults. The results suggest that infants are equipped with remarkable abilities for parsing linguistic input. They are able to identify word-like constituents in finite speech based on predictive syllable relationships. They learn constraints on grammatical word order. They also exhibit rudimentary abstraction abilities, as reflected in their recognition of familiar patterns in novel vocabulary. Finally, they must ultimately discover that the ordering of words in sentences is determined at a more abstract level by dependencies among syntactic categories. We have some understanding of how adults acquire such dependencies, monitoring similar research with younger learners. How does our growing understanding of infant learning abilities bear on the highly constrained language learner described in the introduction? We can identify at least three ways. First, all of the artificial-language-learning studies discussed have examined infants' sensitivity to linguistic form in the absence of semantic content. In so far as these studies are tapping sensitivities used in real-language acquisition, they challenge many accounts in which language development is driven by a mapping between meaning and form. This is not to say that learners do not ultimately need to map the syntactic forms they encode during infancy onto meaning. Obviously they do. However, the fact that infants are able to acquire certain aspects of form prior to acquiring the meaning of these forms changes the nature of the language acquisition problem in a fundamental way.

A second implication of the research on infant artificial-language learning concerns the specificity of the constraints on the learner. On many accounts, these constraints have been contrasted as being language-specific, such that for every aspect of language to be acquired, the child is born with a specific constraint or parameter that guides him/her to the correct representation. Data showing that infants can use transitional probabilities to segment grammatical tone sequences contrasts with this view, suggesting that they apply statistical learning to linguistic and non-linguistic stimuli alike. The application of statistical sensitivity to the problem of word segmentation is admirably far from the constraints discussed by linguistic nativists (involving such language-specific notions as whether or not declarative sentences in a particular language must have an overt subject). Nevertheless, the hypothesis that language (although a specialized human cognitive domain), can be acquired via general-purpose learning mechanisms, is one likely to be investigated with increasing vigor in the next decade.

A third implication of both the infant artificial-language-learning studies reviewed here and the myriad studies of infant language perception preceding them concerns the relevance of children's early utterances as evidence for theories of language acquisition. One of the key observations of linguistic nativism involves errors that children do not make. As noted earlier, children never erroneously transform a statement like 'The man who is tall is Sam' into a question like 'Is the man who tall is Sam?' The lack of such errors, along with logical arguments concerning the poverty of the stimulus, have been taken as evidence that children never consider rules based solely on linear order in sentences. Although researchers have begun to address the question of how a statistical learner might begin to negotiate impoverished input (Box 2), it is equally important to note that if the studies of infants' early linguistic abilities tell us anything, it is that they have become sensitive to many aspects of linguistic form a year or more before they ever begin to produce multiword speech. This is not to say that all of language is acquired by the age of 12 months. However, if infant language-perception studies have one theme, it is in demonstrating the extremely complex (and often contrasting) relationship between aspects of their native language infants and young children have exactly discarded and those they actually produce. Thus, we must exercise caution in interpreting children's early utterances as evidence for or against the linguistic representations they do and do not entertain.

A final comment is in order. Given the vast differences in artificial grammars and natural language, how do we ensure that the learning observed is representative of language learning in the real world? First, in using this approach it is important to design experiments capturing key linguistic phenomena. If we can isolate a phenomenon of interest experimentally, we can go on to test it using a wide range of manipulations, where, presumably, such manipulations are driven by our knowledge of natural language acquisition. For instance, the finding that 18-month-olds, but not 15-month-olds track grammatical word order, suggests that children never consider rules based solely on linear order in sentences. Although researchers have begun to address the question of how a statistical learner might begin to negotiate impoverished input (Box 2), it is equally important to note that if the studies of infants' early linguistic abilities tell us anything, it is that they have become sensitive to many aspects of linguistic form a year or more before they ever begin to produce multiword speech. This is not to say that all of language is acquired by the age of 12 months. However, if infant language-perception studies have one theme, it is in demonstrating the extremely complex (and often contrasting) relationship between aspects of their native language infants and young children have exactly discarded and those they actually produce. Thus, we must exercise caution in interpreting children's early utterances as evidence for or against the linguistic representations they do and do not entertain.

Outstanding questions

- Which aspects of language acquisition are acquired and which are innate?
- How much knowledge is built into the initial system? What is the dynamic between innate and environmental factors? How does this dynamic change over the course of development?
- Do learning mechanisms develop?
- How do domain-general and domain-specific mechanisms factor into language learning?
- Do these mechanisms operate exclusively by means of rules or associations, or do we make use of both symbolic and associative mechanisms?
- To what extent are the mechanisms identified in artificial-language studies the same as those used in acquiring natural language?
References


Erratum

In the Opinion article by M. Tomasello in the April issue of Trends in Cognitive Sciences (Vol. 4, No. 4, pp. 156–163), Table 1 on p. 160 was printed with two errors. In the left-hand column of the table, instead of “Rel: 43” and “Rel: 44” it should read “Lewis and Tomasello (unpublished data)” and “Children and Tomasello (unpublished data)”, respectively.

We apologize to the author and to readers for this oversight.

Pit: 513584-66110010104181-9