

Statistical Language Learning: Mechanisms and Constraints

Jenny R. Saffran¹

Department of Psychology and Waisman Center, University of Wisconsin–Madison, Madison, Wisconsin

Abstract

What types of mechanisms underlie the acquisition of human language? Recent evidence suggests that learners, including infants, can use statistical properties of linguistic input to discover structure, including sound patterns, words, and the beginnings of grammar. These abilities appear to be both powerful and constrained, such that some statistical patterns are more readily detected and used than others. Implications for the structure of human languages are discussed.

Keywords

language acquisition; statistical learning; infants

Imagine that you are faced with the following challenge: You must discover the underlying structure of an immense system that contains tens of thousands of pieces, all generated by combining a small set of elements in various ways. These pieces, in turn, can be combined in an infinite number of ways, although only a subset of those combinations is actually correct. However, the subset that is correct is itself infinite. Somehow you must rapidly figure out the structure of this system so that you can use it appropriately early in your childhood.

This system, of course, is human language. The elements are the sounds of language, and the larger pieces are the words, which in turn combine to form sentences. Given the richness and complexity of lan-

guage, it seems improbable that children could ever discern its structure. The process of acquiring such a system is likely to be nearly as complex as the system itself, so it is not surprising that the mechanisms underlying language acquisition are a matter of long-standing debate. One of the central focuses of this debate concerns the innate and environmental contributions to the language-acquisition process, and the degree to which these components draw on information and abilities that are also relevant to other domains of learning.

In particular, there is a fundamental tension between theories of language acquisition in which learning plays a central role and theories in which learning is relegated to the sidelines. A strength of learning-oriented theories is that they exploit the growing wealth of evidence suggesting that young humans possess powerful learning mechanisms. For example, infants can rapidly capitalize on the statistical properties of their language environments, including the distributions of sounds in words and the orders of word types in sentences, to discover important components of language structure. Infants can track such statistics, for example, to discover speech categories (e.g., native-language consonants; see, e.g., Maye, Werker, & Gerken, 2002), word boundaries (e.g., Saffran, Aslin, & Newport, 1996), and rudimentary syntax (e.g., Gomez & Gerken, 1999; Saffran & Wilson, 2003).

However, theories of language acquisition in which learning plays a central role are vulnerable to a number of criticisms. One of the

most important arguments against learning-oriented theories is that such accounts seem at odds with one of the central observations about human languages. The linguistic systems of the world, despite surface differences, share deep similarities, and vary in non-arbitrary ways. Theories of language acquisition that focus primarily on preexisting knowledge of language do provide an elegant explanation for cross-linguistic similarities. Such theories, which are exemplified by the seminal work of Noam Chomsky, suggest that linguistic universals are prespecified in the child's linguistic endowment, and do not require learning. Such accounts generate predictions about the types of patterns that should be observed cross-linguistically, and lead to important claims regarding the evolution of a language capacity that includes innate knowledge of this kind (e.g., Pinker & Bloom, 1990).

Can learning-oriented theories also account for the existence of language universals? The answer to this question is the object of current research. The *constrained statistical learning framework* suggests that learning is central to language acquisition, and that the specific nature of language learning explains similarities across languages. The crucial point is that learning is constrained; learners are not open-minded, and calculate some statistics more readily than others. Of particular interest are those constraints on learning that correspond to cross-linguistic similarities (e.g., Newport & Aslin, 2000). According to this framework, the similarities across languages are indeed nonaccidental, as suggested by the Chomskian framework—but they are not the result of innate linguistic knowledge. Instead, human languages have been shaped by human learning mechanisms (along with constraints on human perception, processing, and speech production), and aspects of language that enhance

learnability are more likely to persist in linguistic structure than those that do not. Thus, according to this view, the similarities across languages are not due to innate knowledge, as is traditionally claimed, but rather are the result of constraints on learning. Further, if human languages were (and continue to be) shaped by constraints on human learning mechanisms, it seems likely that these mechanisms and their constraints were not tailored solely for language acquisition. Instead, learning in nonlinguistic domains should be similarly constrained, as seems to be the case.

A better understanding of these constraints may lead to new connections between theories focused on nature and theories focused on nurture. Constrained learning mechanisms require both particular experiences to drive learning and pre-existing structures to capture and manipulate those experiences.

LEARNING THE SOUNDS OF WORDS

In order to investigate the nature of infants' learning mechanisms, my colleagues and I began by studying an aspect of language that we knew must certainly be learned: word segmentation, or the boundaries between words in fluent speech. This

is a challenging problem for infants acquiring their first language, for speakers do not mark word boundaries with pauses, as shown in Figure 1. Instead, infants must determine where one word ends and the next begins without access to obvious acoustic cues. This process requires learning because children cannot innately know that, for example, *pretty* and *baby* are words, whereas *tyba* (spanning the boundary between *pretty* and *baby*) is not.

One source of information that may contribute to the discovery of word boundaries is the statistical structure of the language in the infant's environment. In English, the syllable *pre* precedes a small set of syllables, including *ty*, *tend*, and *cedes*; in the stream of speech, the probability that *pre* is followed by *ty* is thus quite high (roughly 80% in speech to young infants). However, because the syllable *ty* occurs word finally, it can be followed by any syllable that can begin an English word. Thus, the probability that *ty* is followed by *ba*, as in *pretty baby*, is extremely low (roughly 0.03% in speech to young infants). This difference in sequential probabilities is a clue that *pretty* is a word, and *tyba* is not. More generally, given the statistical properties of the input language, the ability to track sequential probabilities would be an extremely useful tool for infant learners.

To explore whether humans can use statistical learning to segment

words, we exposed adults, first graders, and 8-month-olds to spoken nonsense languages in which the only cues to word boundaries were the statistical properties of the syllable sequences (e.g., Saffran et al., 1996). Listeners briefly heard a continuous sequence of syllables containing multisyllabic words from one of the languages (e.g., *golabupabikutitubabupugolabubabupu. . .*). We then tested our participants to determine whether they could discriminate the words from the language from sequences spanning word boundaries. For example, we compared performance on words like *golabu* and *pabiku* with performance on sequences like *bupabi*, which spanned the boundary between words. To succeed at this task, listeners would have had to track the statistical properties of the input. Our results confirmed that human learners, including infants, can indeed use statistics to find word boundaries. Moreover, this ability is not confined to humans: Cotton-top tamarins, a New World monkey species, can also track statistics to discover word boundaries (Hauser, Newport, & Aslin, 2001).

One question immediately raised by these results is the degree to which statistical learning is limited to languagelike stimuli. A growing body of results suggests that sequential statistical learning is quite general. For example, infants can track sequences of tones, discovering "tone-word boundaries" via statistical cues (e.g., Saffran, Johnson, Aslin, & Newport, 1999), and can learn statistically defined visual patterns (e.g., Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002); work in progress is extending these results to the domain of events in human action sequences.

Given that the ability to discover units via their statistical coherence is not confined to language (or to humans), one might wonder whether the statistical learning results actually pertain to language at all. That

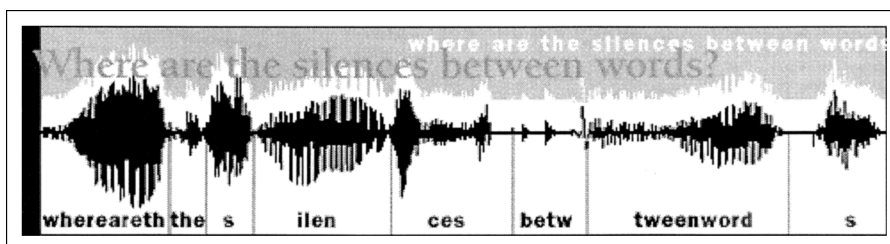


Fig. 1. A speech waveform of the sentence "Where are the silences between words?" The height of the bars indicates loudness, and the x-axis is time. This example illustrates the lack of consistent silences between word boundaries in fluent speech. The vertical gray lines represent quiet points in the speech stream, some of which do not correspond to word boundaries. Some sounds are represented twice in the transcription below the waveform because of their continued persistence over time.

is, do infants actually use statistical learning mechanisms in real-world language acquisition? One way to address this question is to ask what infants are actually learning in our segmentation task. Are they learning statistics? Or are they using statistics to learn language? Our results suggest that when infants being raised in English-speaking environments have segmented the sound strings, they treat these nonsensical patterns as English words (Saffran, 2001b). Statistical language learning in the laboratory thus appears to be integrated with other aspects of language acquisition. Related results suggest that 12-month-olds can first segment novel words and then discover syntactic regularities relating the new words—all within the same set of input. This would not be possible if the infants formed mental representations only of the sequential probabilities relating individual syllables, and no word-level representations (Saffran & Wilson, 2003). These findings point to a constraint on statistical language learning: The mental representations produced by this process are not just sets of syllables linked by statistics, but new units that are available to serve as the input to subsequent learning processes.

Similarly, it is possible to examine constraints on learning that might affect the acquisition of the sound structure of human languages. The types of sound patterns that infants learn most readily may be more prevalent in languages than are sound patterns that are not learnable by infants. We tested this hypothesis by asking whether infants find some phonotactic regularities (restrictions on where particular sounds can occur; e.g., /fs/ can occur at the end, but not the beginning, of syllables in English) easier to acquire than others (Saffran & Thiessen, 2003). The results suggest that infants readily acquire novel regularities that are consistent with the types of patterns found in the

world's languages, but fail to learn regularities that are inconsistent with natural language structure. For example, infants rapidly learn new phonotactic regularities involving generalizations across sounds that share a phonetic feature, while failing to learn regularities that disregard such features. Thus, it is easier for infants to learn a set of patterns that group together /p/, /t/, and /k/, which are all voiceless, and that group together /b/, /d/, and /g/, which are all voiced, than to learn a pattern that groups together /d/, /p/, and /k/, but does not apply to /t/.² Studies of this sort may provide explanations for why languages show the types of sound patterning that they do; sound structures that are hard for infants to learn may be unlikely to recur across the languages of the world.

STATISTICAL LEARNING AND SYNTAX

Issues about learning versus innate knowledge are most prominent in the area of syntax. How could learning-oriented theories account for the acquisition of abstract structure (e.g., phrase boundaries) not obviously mirrored in the surface statistics of the input? Unlike accounts centered on innate linguistic knowledge, most learning-oriented theories do not provide a transparent explanation for the ubiquity of particular structures cross-linguistically. One approach to these issues is to ask whether some nearly universal structural aspects of human languages may result from constraints on human learning (e.g., Morgan, Meier, & Newport, 1987). To test this hypothesis, we asked whether one such aspect of syntax, phrase structure (groupings of types of words together into subunits, such as noun phrases and verb phrases), results from a constraint on learning: Do humans learn sequen-

tial structures better when they are organized into subunits such as phrases than when they are not? We identified a statistical cue to phrasal units, predictive dependencies (e.g., the presence of a word like *the* or *a* predicts a noun somewhere downstream; the presence of a preposition predicts a noun phrase somewhere downstream), and determined that learners can use this kind of cue to locate phrase boundaries (Saffran, 2001a).

In a direct test of the theory that predictive dependencies enhance learnability, we compared the acquisition of two nonsense languages, one with predictive dependencies as a cue to phrase structure, and one lacking predictive dependencies (e.g., words like *the* could occur either with or without a noun, and a noun could occur either with or without words like *the*; neither type of word predicted the presence of the other). We found better language learning in listeners exposed to languages containing predictive dependencies than in listeners exposed to languages lacking predictive dependencies (Saffran, 2002). Interestingly, the same constraint on learning emerged in tasks using nonlinguistic materials (e.g., computer alert sounds and simultaneously presented shape arrays). These results support the claim that learning mechanisms not specifically designed for language learning may have shaped the structure of human languages.

DIRECTIONS FOR FUTURE RESEARCH

Results to date demonstrate that human language learners possess powerful statistical learning capacities. These mechanisms are constrained at multiple levels; there are limits on what information serves as input, which computations are performed over that input, and the

structure of the representations that emerge as output. To more fully understand the contribution of statistical learning to language acquisition, it is necessary to assess the degree to which statistical learning provides explanatory power given the complexities of the acquisition process.

For example, how does statistical learning interact with other aspects of language acquisition? One way we are addressing this question is by investigating how infants weight statistical cues relative to other cues to word segmentation early in life. The results of such studies provide an important window into the ways in which statistical learning may help infant learners to determine the relevance of the many cues inherent in language input. Similarly, we are studying how statistics meet up with meaning in the world (e.g., are statistically defined "words" easier to learn as labels for novel objects than sound sequences spanning word boundaries?), and how infants in bilingual environments cope with multiple sets of statistics. Studying the intersection between statistical learning and the rest of language learning may provide new insights into how various nonstatistical aspects of language are acquired. Moreover, a clearer picture of the learning mechanisms used successfully by typical language learners may increase researchers' understanding of the types of processes that go awry when children do not acquire language as readily as their peers.

It is also critical to determine which statistics are available to young learners and whether those statistics are actually relevant to natural language structure. Researchers do not agree on the role that statistical learning should play in acquisition theories. For example, they disagree about when learning is best described as statistically based as opposed to rule based (i.e., utilizing mechanisms that operate over algebraic variables to discover

abstract knowledge), and about whether learning can still be considered statistical when the input to learning is abstract. Debates regarding the proper place for statistical learning in theories of language acquisition cannot be resolved in advance of the data. For example, although one can distinguish between statistical versus rule-based learning mechanisms, and statistical versus rule-based knowledge, the data are not yet available to determine whether statistical learning itself renders rule-based knowledge structures, and whether abstract knowledge can be probabilistic. Significant empirical advances will be required to disentangle these and other competing theoretical distinctions.

Finally, cross-species investigations may be particularly informative with respect to the relationship between statistical learning and human language. Current research is identifying species differences in the deployment of statistical learning mechanisms (e.g., Newport & Aslin, 2000). To the extent that nonhumans and humans track different statistics, or track statistics over different perceptual units, learning mechanisms that do not initially appear to be human-specific may actually render human-specific outcomes. Alternatively, the overlap between the learning mechanisms available across species may suggest that differences in statistical learning cannot account for cross-species differences in language-learning capacities.

CONCLUSION

It is clear that human language is a system of mind-boggling complexity. At the same time, the use of statistical cues may help learners to discover some of the patterns lurking in language input. To what extent might the kinds of statistical

patterns accessible to human learners help in disentangling the complexities of this system? Although the answer to this question remains unknown, it is possible that a combination of inherent constraints on the types of patterns acquired by learners, and the use of output from one level of learning as input to the next, may help to explain why something so complex is mastered readily by the human mind. Human learning mechanisms may themselves have played a prominent role in shaping the structure of human languages.

Recommended Reading

- Gómez, R.L., & Gerken, L.A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.
- Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.
- Pena, M., Bonatti, L.L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.
- Seidenberg, M.S., MacDonald, M.C., & Saffran, J.R. (2002). Does grammar start where statistics stop? *Science, 298*, 553–554.

Acknowledgments—The preparation of this manuscript was supported by grants from the National Institutes of Health (HD37466) and National Science Foundation (BCS-9983630). I thank Martha Alibali, Erin McMullen, Seth Pollak, Erik Thiessen, and Kim Zinski for comments on a previous version of this manuscript.

Notes

1. Address correspondence to Jenny R. Saffran, Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706; e-mail: jsaffran@wisc.edu.

2. Voicing refers to the timing of vibration of the vocal cords. Compared with voiceless consonants, voiced consonants have a shorter lag time between the initial noise burst of the consonant and the subsequent vocal cord vibrations.

References

- Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA*, 99, 15822–15826.
- Gomez, R.L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- Hauser, M., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B41–B52.
- Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition*, 83, B35–B42.
- Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- Morgan, J.L., Meier, R.P., & Newport, E.L. (1987). Structural packaging in the input to language learning: Contributions of intonational and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.
- Newport, E.L., & Aslin, R.N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S.C. Howell, S.A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Boston University Conference on Language Development* (pp. 1–21). Somerville, MA: Cascadia Press.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707–784.
- Saffran, J.R. (2001a). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.
- Saffran, J.R. (2001b). Words in a sea of sounds: The output of statistical learning. *Cognition*, 81, 149–169.
- Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172–196.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Saffran, J.R., & Thiessen, E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484–494.
- Saffran, J.R., & Wilson, D.P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy*, 4, 273–284.

The Origins of Pictorial Competence

Judy S. DeLoache,¹ Sophia L. Pierroutsakos, and David H. Uttal

Department of Psychology, University of Virginia, Charlottesville, Virginia (J.S.D.);
Department of Psychology, Furman University, Greenville, South Carolina (S.L.P.);
and Department of Psychology, Northwestern University, Evanston, Illinois
(D.H.U.)

Abstract

Pictorial competence, which refers to the many factors involved in perceiving, interpreting, understanding, and using pictures, develops gradually over the first few years of life. Although experience is not required for accurate perception of pictures, it is necessary for understanding the nature of pictures. Infants initially respond to depicted objects as if they were real objects, and toddlers are remarkably insensitive to picture orientation. Only gradually do young children figure out the nature of pictures and how they are used.

Keywords

symbolic development; picture perception

As philosophers, new and old, have emphasized, humans are “the

symbolic species” (Deacon, 1997), and symbolization is the “most characteristic mental trait of [humans]” (Langer, 1942, p. 72). Just as the emergence of the symbolic capacity in the course of evolution irrevocably transformed the human species, so too does the development of symbolic functioning transform young children. The capacity for symbolization vastly expands their intellectual horizons, liberating them from the constraints of time and space and enabling them to acquire information about reality without directly experiencing it.

All children growing up anywhere in the world must master a wide variety of symbol systems and symbolic artifacts for full participation in their society. Our research has focused on how young children begin to understand and exploit the informational potential of various symbolic objects, including models, maps, and pictures.

We define a symbolic artifact as something that someone intends to stand for something other than itself (DeLoache, 1995). Thus, virtually anything can serve as a symbol, and virtually any concept that one has can be symbolized, but the symbol is always different in some way from that which it represents. What makes something symbolic is human intention; an entity becomes a symbol only as the result of a person using it to denote or refer to something.

THE CHALLENGE OF DUAL REPRESENTATION

Although mastering symbols is a universal task, it is not an easy one. A formidable challenge to young children in developing competence with symbols stems from the inherently dual nature of symbols; every symbolic artifact is an object in and of itself, and at the same time it also stands for something other than itself. To understand and use a symbol, *dual representation* is necessary—one must mentally represent both facets of the symbol’s dual reality, both its concrete characteristics and its abstract relation to what it stands for