Statistical language learning and language acquisition

Katharine Graf Estes
University of California, Davis

During the first year of life, infants learn a great deal about the language in which they are immersed. The precocious nature of infants' language acquisition skills has led to the investigation of mechanisms that support early learning. Statistical learning is gaining attention as a potentially powerful learning mechanism that may allow infants to break into the structure of language well before they produce their first words. Infants can use the process of statistical learning to detect regularities in linguistic input, including phoneme categories, word boundaries, and simple syntactic patterns (reviewed in Gomez & Gerken, 2000; Saffran, 2003). Following the initial demonstrations of infants' ability to track patterns of co-occurrences in linguistic (and non-linguistic) input, recent investigations of statistical learning have addressed how infants might apply this learning mechanism to fundamental tasks of language acquisition, such as learning of vocabulary and syntax.

Recent work investigating the connection between statistical learning and word learning suggests that the ability to track distributions of syllable co-occurrences in fluent speech may provide infants with statistically segmented sound sequences that are available to link to meaning (Graf Estes, Evans, Alibali, Saffran, 2007). Statistical learning may also support word learning by providing infants with information about likely and unlikely phonotactic patterns in their native language, thus readying likely sound sequences to be linked to meanings. Statistical learning may play a role in early syntactic acquisition as well. For example, infants can use the output of statistical word segmentation (i.e., newly segmented words) to discover the permissible orderings of words in utterances (Saffran & Wilson, 2003; see also Gomez & Gerken, 1999).

However, a growing body of studies suggests that learners are not open to tracking all possible statistics. A learning mechanism that attempts to track all possible correlations between elements, and that weighs all correlations equally, will quickly become bogged down in computations. Rather, it seems that some regularities in sound and word combinations are calculated more readily than others (e.g., Newport & Aslin, 2004; Saffran, 2001; Saffran & Thiessen, 2003). Further, the patterns that infants, children, and adults readily discover in experimentally-controlled artificial languages tend to correspond to structures that appear in natural languages. Conversely, patterns that do not commonly occur across languages also tend to be difficult to learn in artificial languages. Understanding the constraints on the range of sound-level and word-level regularities that learners compute will promote a more precise characterization of how statistical learning supports learning of vocabulary and syntax. More broadly, evidence for constraints on statistical learning suggest language evolution may be shaped by the nature of human learning mechanisms.

General outline, with examples of work to be discussed

1. Infants as impressive statistical learners.
   (e.g., Saffran, Aslin, & Newport)

2. The application of statistical learning to word learning.
   (Graf Estes, Evans, Alibali, & Evans, 2007; Graf Estes (2007); Graf Estes & Saffran, in preparation; Swingley, 2003)

3. The application of statistical learning to grammar learning.
   (Saffran & Wilson, 2003; Mintz, 2003)

4. Constraints on statistical learning of sounds (Saffran & Thiessen, 2003) and grammars (Saffran, 2001; Newport & Aslin, 2000; 2004).

5. Discussion of statistical learning constraints and possible connections to language evolution. Languages may have been shaped by the nature of human learning mechanisms—avoiding constructions that are difficult to learn and promoting constructions that fit how infants and children learn, and how adults process language.

References

Graf Estes, K., Evans, J. L., Alibali, M., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18 (2),* 254-260.

Graf Estes, K. (2007, June). Is there anything wordlike about statistical word segmentation? Paper presented at the Workshop on Current Issues in Language Acquisition: Artificial languages and statistical learning, Calgary, Alberta Canada.

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

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

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

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, and T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*, Somerville, MA: Cascadilla Press, 1-21.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Saffran, J. R., Aslin, R. N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J.R., & Thiessen E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*, 1926-1928.

Saffran, J. R., & Wilson D.P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4,* 273-284.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86-132.