Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 105 (2007) 247–299
COGNITION
www.elsevier.com/locate/COGNIT

Rapid learning of syllable classes from a perceptually continuous speech stream ☆, ☆☆

Ansgar D. Endress [a,b,*], Luca L. Bonatti [b,c]

[a] *Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-ENS-CNRS, Paris, France*
[b] *International School for Advanced Studies, Trieste, Italy*
[c] *University of Paris VIII at Saint Denis, France*

Received 12 July 2005; revised 14 August 2006; accepted 9 September 2006

Abstract

To learn a language, speakers must learn its words and rules from fluent speech; in particular, they must learn dependencies among linguistic classes. We show that when familiarized with a short artificial, subliminally bracketed stream, participants can learn relations about the structure of its words, which specify the classes of syllables occurring in first and last word positions. By studying the effect of familiarization length, we compared the general predictions of associative theories of learning and those of models postulating separate mechanisms for quickly extracting the word structure and for tracking the syllable distribution in the stream. As predicted by the dual-mechanism model, the preference for structurally correct items was negatively correlated with the familiarization length. This result is difficult to explain by purely associative schemes; an extensive set of neural network simulations confirmed this difficulty. Still, we show that powerful statistical computations operating on the stream are available to our participants, as they are sensitive to co-occurrence statistics among non-adjacent syllables. We suggest that different learning mechanisms analyze speech on-line: A rapid mechanism extracting structural information about the stream, and a slower mechanism detecting statistical regularities among the items occurring in it.

☆ This manuscript was accepted under the editorship of Jacques Mehler.
☆☆ This research was supported by an Allocation de Moniteur Polytechnicien fellowship to ADE and by the HFSP Grant RGP 68/2002. We are grateful to Ghislaine Dehaene-Lambertz, Jacques Mehler, Marina Nespor, Marcela Peña, Alessandro Treves, Judit Gervain and Valerie Lesk for discussions and comments on an earlier version of this manuscript.
* Corresponding author. Tel.: +39 040 3787 603; fax: +39 040 3787 615.
E-mail address: ansgar.endress@polytechnique.org (A.D. Endress).

doi:10.1016/j.cognition.2006.09.010

Keywords: Artificial grammar learning; Language acquisition; Dual mechanism models; Symbol manipulation

1. The multi-faced problem of language learning

Infants learn their native language fast and accurately despite the complex computations involved in language learning. One of the challenges they face is to learn the words of a language. Although adults perceive speech as a discrete sequence of words, no reliable acoustic cues indicate word boundaries in a speech stream. Therefore, to construct a lexicon, infants must first individuate the sound stretches in the continuous speech signal that form words.

It has long been known that sensitivity to statistical cues in the speech signal, in particular to transition probabilities (TPs), could in principle solve this problem (e.g., Harris, 1955; Hayes & Clark, 1970, but see Yang, 2004).[1] Recent research has shown that adults and infants can indeed segment a continuous speech stream when the only cue to word boundaries is that TPs are high within words and low between words (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996; Saffran, Aslin, & Newport, 1996).

Sensitivity to statistical cues such as TPs might help infants to construct a lexicon. However, the lexicon is not a mere list of word-meaning pairs, but also contains a vast amount of structural information. For example, the lexical entry of the verb "give" must contain a specification of the relationship between the specific lexical entry (GIVE) and the arguments it takes (the giver, the given and the goal). Not only is such structural information commonly assumed in linguistic theories (e.g., Cook & Newson, 1996), but it also appears to be a requirement for word learning in the first place (Gillette, Gleitman, Gleitman, & Lederer, 1999; Landau & Gleitman, 1985). Structural information also plays a prominent role in morphology and syntax. Both require not only to individuate single words, but also to represent structural relations between sub-lexical morphemes or syntactic word classes. Since infants will eventually be able to use the lexicon productively, they cannot just memorize the words and sentences they have heard, but have to generalize grammatical and morphological regularities.

In spite of the importance of structural information in language, little is known about the processes underlying its acquisition. Indeed, it is still a matter of much debate whether learners do extract structural information, or whether they only track associations among items. Even in a well-studied case such as that of past tense formation, most studies either focus on adult speakers with years of linguistic experience, or try to model speakers' production by computer simulations. Hence, these

[1] TP is the conditional probability of encountering a syllable after having encountered another syllable. After having encountered the syllable /don/, there is a high probability of encountering the syllable /key/ because "donkey" is a word. More formally, conditional probabilities like $P(\sigma_{i+1} = /key/ \mid \sigma_i = /don/)$ are high within words, and low between words (σ denotes syllables in a speech stream).