Journal of Memory and Language 72 (2014) 49–58

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Statistically coherent labels facilitate categorization in 8-month-olds

Lucy C. Erickson [a,*], Erik D. Thiessen [a], Katharine Graf Estes [b]

[a] Carnegie Mellon University, United States
[b] University of California, Davis, United States

ARTICLE INFO

Article history:
Received 16 August 2013
revision received 22 November 2013
Available online 29 January 2014

Keywords:
Statistical learning
Word segmentation
Labels
Categorization

ABSTRACT

There is considerable evidence that infants can segment speech using syllable co-occurrence probabilities; however, relatively less is known about the nature of the representations formed during this process. The present studies tested the prediction that statistically segmented items should exhibit a specific property of real words, namely, these items should have a facilitative effect on infant categorization. During the segmentation phase, eight-month-old infants listened to a fluent speech stream that contained statistical word boundary cues. Infants were then tested on their ability to categorize drawings of an unfamiliar category when category exemplars were paired with either high-probability or low-probability labels from the segmentation phase. Infants who heard high-probability labels showed evidence of categorization. In contrast, infants who heard low-probability labels did not. A follow up experiment revealed that this effect was due to facilitation for high-probability words rather than inhibition for low-probability items. These results fit with theoretical accounts that suggest that infants treat statistically segmented units as potential words.

© 2014 Elsevier Inc. All rights reserved.

Introduction

A critical task that infants face in the first years of life is to acquire a lexicon. This task is challenging because infants hear relatively few words in isolation (Brent & Siskind, 2001), and because word boundaries in fluent speech are marked by a combination of imperfect cues rather than a single infallible cue (Cole & Jakimik, 1980). Consequently, the question of how infants accomplish word segmentation has generated a large body of research (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999; Saffran, Aslin, & Newport, 1996; for a review, see Jusczyk, 1999). One possibility is that infants rely on a cross-linguistically consistent characteristic of words, namely, that syllables that comprise words cluster together with greater reliability than syllables that occur incidentally across word boundaries (Harris, 1955; Hayes & Clark, 1970). Thus, a mechanism that infants may use to segment their first words from continuous speech is statistical word segmentation, the process of segmenting units from speech on the basis of syllable co-occurrence probabilities.

Statistical learning is a powerful learning mechanism that is available from birth to adulthood (Graf Estes, Evans, Alibali, & Saffran, 2007; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). In statistical word segmentation experiments, learners require only a brief familiarization with a fluent speech passage to track syllable patterns that identify words (high probability clusters) versus across-word sequences (low probability clusters). Moreover, because this mechanism does not depend on language-specific knowledge, it may be of particular use in infants' initial discovery of word forms. As such, it may be the case that statistical learning plays an important role in early lexical acquisition (e.g.,

* Corresponding author. Fax: +1 412 286 2798.
E-mail address: lerickso@andrew.cmu.edu (L.C. Erickson).

0749-596X/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2014.01.002

Saffran & Thiessen, 2007; Thiessen & Saffran, 2007; Werker & Curtin, 2005). Consistent with this possibility, statistical learning appears to be related to various aspects of lexical development (e.g., Evans, Saffran, & Robe-Torres, 2009; Romberg & Saffran, 2010).

However, there have been objections to the claim that statistical learning contributes to language development. One centers on the fact that most studies of statistical learning have been conducted in laboratory settings with artificial stimuli. This has raised questions about whether the learning processes that are found in the laboratory are the same as those involved in developing a lexicon in infants' natural environments (e.g., Graf Estes, 2012; Johnson & Tyler, 2010). A related objection is that the representations that emerge from statistical learning are not linguistic or lexical in nature. For example, Endress and Mehler (2009) assert that statistical word segmentation does not result in word-like representations. Rather, they have proposed that although learners are able to compute transitional probabilities between syllables, there is no evidence that "the items with stronger [transitional probabilities] are represented as actual word-like units, or even that they have been extracted" (p. 352). Instead, they have argued that statistical learning produces chaining memories that correspond to syllable transitions, but does not result in the storage of an integrated percept (but see Perruchet & Poulin-Charronnat, 2012). These issues are problematic for accounts that propose that infants use statistical segmentation to discover their first word forms. If infants do use statistical segmentation for lexical acquisition, they should treat the output of statistical learning as potential word forms. The goal of the current research was to test whether the product of statistical learning exhibits an established property of real words, which exert an influence on infant object categorization.

To address questions about the nature of the representations emerging from statistical learning, researchers have attempted to determine whether they exhibit the properties of real words. One approach to this question has been to assess whether computational models of statistical learning give rise to representations that are similar to those thought to exist in the lexicon. For example, chunking models of statistical learning give rise to the kinds of unitary representations that are thought to characterize lexical items (e.g., Perruchet & Vinter, 1998; Giroux & Rey, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008; Thiessen, Kronstein, & Hufnagle, 2013). A complementary approach to this question is to use behavioral methods to probe the representations resulting from infant statistical learning. In one such study, Saffran (2001) tested whether preferences for statistical nonsense words parsed from fluent speech (e.g., *tibudo*) differed as a function of the lexical context. These nonsense words were either embedded into a highly familiar English sentence frame (e.g., "I like my tibudo") or one that was a nonsense sentence matched on several dimensions (e.g., "Zy fike ny tibudo"). Eight-month-old infants listened longer to sentences that included statistically coherent words from the speech stream than to sentences containing sequences that crossed word boundaries. However, this preference only held when the words were presented in the context of a real English sentence. There was no difference in listening times for infants who were exposed to nonsense word frames, which indicates that they only treated these items differently in meaningful linguistic contexts.

Graf Estes et al. (2007) used a word learning paradigm to investigate whether the process of statistically segmenting words from fluent speech is related to the process of mapping meanings to labels. The authors found that 17-month-old infants were able to map labels to objects when those labels comprised syllable sequences with high internal probabilities, based on prior statistical segmentation experience. In contrast, infants did not learn the mapping when the labels comprised familiar sequences with low internal probabilities. This indicates that the process of segmenting words from fluent speech is intimately linked to word learning. Because infants had already segmented these labels and stored them in memory, cognitive resources that might otherwise have been devoted to learning the word form were free to learn the mapping between the label and its referent. Taken together, Saffran (2001) and Graf Estes et al.'s (2007) studies suggest that infants treat high probability items, and thus the output of statistical learning, as possible native language words rather than merely sound sequences with high internal probabilities but no linguistic relevance.

We hypothesized that if the output of statistical learning is indeed word-like, it should exhibit additional properties of real words. One such property relates to the relationship between linguistic labels and infant categorization. A body of research has explored the effects of linguistic and nonlinguistic labels on infant object categorization (e.g., Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Fulkerson & Haaf, 2003; Namy & Waxman, 1998). In this literature, infants are familiarized with exemplars of novel object categories (e.g., rabbits) paired with a consistent label (e.g., "A *toma*" or "Look at the *toma*"). They are subsequently tested on their knowledge of the category as a whole as indexed by their ability to discriminate a novel exemplar of the familiar category (i.e., a rabbit) from a novel exemplar of a new category (e.g., a pig). Research using this procedure has found that the types of symbols that are conducive to category formation tend to be symbols that are used to refer to objects in infants' real environments (Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007; Namy & Waxman, 1998; Woodward & Hoyne, 1999). For example, infants readily form an object category (e.g., dinosaurs) when exemplars of that category are consistently accompanied by a linguistic label; when the exemplars are accompanied by tones they do not exhibit categorization (e.g., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007).

The relationship between label identity and categorization performance is not static over development. Instead, the range of signals that permit categorization appears to begin relatively broadly and to narrow over the course of development, as infants become attuned to their native language. A recent study demonstrated that 3- to 4-month-old infants categorize in the presence of nonhuman primate vocalizations (Ferry, Hespos, & Waxman, 2013), similar to their performance with linguistic naming

Download English Version:

https://daneshyari.com/en/article/931878

Download Persian Version:

https://daneshyari.com/article/931878

Daneshyari.com