Journal of Memory and Language 60 (2009) 351–367

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words

Ansgar D. Endress[a,*], Jacques Mehler[b]

[a] Department of Psychology, Harvard University, William James Hall 984, 33 Kirkland Street, Cambridge, MA 02138, USA
[b] International School for Advanced Studies, Trieste, Italy

ARTICLE INFO

Article history:
Received 22 January 2008
Revision received 12 October 2008
Available online 4 February 2009

Keywords:
Statistical learning
Transition probabilities
Word segmentation
Cue integration
Perceptual cues to word learning

ABSTRACT

Word-segmentation, that is, the extraction of words from fluent speech, is one of the first problems language learners have to master. It is generally believed that statistical processes, in particular those tracking "transitional probabilities" (TPs), are important to word-segmentation. However, there is evidence that word forms are stored in memory formats differing from those that can be constructed from TPs, i.e. in terms of the positions of phonemes and syllables within words. In line with this view, we show that TP-based processes leave learners no more familiar with items heard 600 times than with "phantom-words" not heard at all if the phantom-words have the same statistical structure as the occurring items. Moreover, participants are more familiar with phantom-words than with frequent syllable combinations. In contrast, minimal prosody-like perceptual cues allow learners to recognize actual items. TPs may well signal co-occurring syllables; this, however, does not seem to lead to the extraction of word-like units. We review other, in particular prosodic, cues to word-boundaries which may allow the construction of positional memories while not requiring language-specific knowledge, and suggest that their contributions to word-segmentation need to be reassessed.

Introduction

Speech comes as a continuous signal, with no reliable cues to signal word boundaries. Thus learners have not only to map the words of their native language to their meanings (which is in itself a difficult problem), but first they have to identify the sound stretches corresponding to words. Thus, they need mechanisms that allow them to memorize the phonological forms of the words they encounter in fluent speech. Here we ask what kinds of memory mechanisms they can employ for this purpose. It is generally accepted that statistical computations are well suited for segmenting words from fluent speech, and thus for memorizing phonological word-candidates (e.g., Aslin, Saffran, & Newport, 1998; Cairns, Shillcock, Levy, & Chater, 1997; Elman, 1990; Goodsitt, Morgan, & Kuhl, 1993; Hayes & Clark, 1970; Saffran, 2001b; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Swingley, 2005). However, as reviewed below in more detail, there is evidence, in particular from speech errors, that memory for words in fact appeals to different kinds of memory mechanisms, namely those encoding the *positions* of phonemes or syllables within words. We thus ask whether learners extract word-like units from fluent speech when just the aforementioned statistical cues are given, or whether they require other, possibly prosodic, cues that allow them to construct positional memories. Specifically, we presented participants with continuous speech streams containing statistically defined "words". These words were chosen such that there were statistically matched "phantom-words" that, despite having the same statistical structure as words, never occurred in the speech streams. If statistical cues lead to the extraction of words from fluent speech, participants should know that they have encountered words but not phantom-words during the speech streams. In contrast, if memory for words is positional, participants should fail to prefer words to

* Corresponding author. Fax: +1 617 495 3886.
E-mail address: ansgar.endress@m4x.org (A.D. Endress).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2008.10.003

phantom-words when only statistical information is given. Rather such a preference should arise only once cues are available that lead to the construction of positional memories.

Evidence for co-occurrence statistics as cues to word boundaries

Once they reach a certain age, learners can use many different cues to predict word boundaries (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Cutler & Norris, 1988; Dahan & Brent, 1999; Jusczyk, Cutler, & Redanz, 1993; Mattys & Jusczyk, 2001; Shukla, Nespor, & Mehler, 2007; Suomi, McQueen, & Cutler, 1997; Thiessen & Saffran, 2003; Vroomen, Tuomainen, & de Gelder, 1998). However, many of these cues are language-specific, and thus have to be learned. For instance, if learners assume that strong syllables are word-initial, they will be right in Hungarian but wrong in French (where strong syllables are word-final), and to learn where stress falls in a word, they have to know the words in the first place. Hence, at least initially, language learners need to use cues to word-boundaries that do not require any language-specific knowledge.

Co-occurrence statistics such as transitional probabilities (TPs) among syllables are one such cue that is particularly well-attested. These statistics indicate how likely it is that two syllables will follow each other. More formally, TPs are conditional probabilities of encountering a syllable after having encountered another syllable. Conditional probabilities like $P(\sigma_{i+1} = \text{pet}|\sigma_i = \text{trum})$ (in the word trumpet) are high within words, and low between words (σ denotes a syllable in a speech stream). Dips in TPs may give cues to word boundaries, while high-TP transitions may indicate that words continue. That is, learners may postulate word boundaries between syllables that rarely follow each other. Saffran and collaborators (e.g., Aslin et al., 1998; Saffran et al., 1996) have shown that even young infants can deploy such statistical computations on continuous speech streams. After familiarization with speech streams in which dips in TPs were the only cue to word boundaries, 8-month-old infants were more familiar with items delimited by TP dips than with items that straddle such dips. Even more impressively, after such a familiarization, infants recognize the items delimited by dips in TPs in *new* English sentences pronounced by a *new* speaker (Saffran, 2001b), suggesting that TP-based segmentation procedures may lead infants to extract word-like units.

Results such as these have led to the widespread agreement that co-occurrence statistics are important for segmenting words from speech. Though not thought to be the only cues used for word-segmentation, they are thought to play a particularly prominent role because, unlike other cues, they can be used by infants without any knowledge of the properties of their native language (e.g., Thiessen & Saffran, 2003).[1] Moreover, similar computations have been observed with other auditory and visual stimuli (Fiser & Aslin, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005), and with other mammals (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005). Such computations may thus be domain- and species-general, stressing again the potential importance of such processes for a wide array of cognitive learning situations. Accordingly, some authors have proposed that these processes may be crucial not only for word-learning but also for other, more grammatical aspects of language acquisition (Bates & Elman, 1996; Saffran, 2001a; Thompson & Newport, 2007).

Surprisingly, however, there is no evidence that TP-based computations lead to the extraction of word-candidates. The experiments above have provided numerous demonstrations that participants are more familiar with items with stronger TPs than with items with weaker TPs. This, however, does not imply that the items with stronger TPs are represented as actual word-like units, or even that they have been extracted. For example, one may well find that a piece of cheese is more associated with a glass of wine than with a glass of beer, but this does not imply that the wine/cheese combination is represented as a unit for parsing the visual scene. Likewise, choosing items with stronger TPs (where the syllables have stronger associations) over items with weaker TPs does not imply either that the items with stronger TPs have been extracted as perceptual units.

The distinction between a preference for high-TP items and representing these items as perceptual units is well illustrated in Turk-Browne and Scholl (2009) studies of visual statistical learning. In these experiments, participants saw a continuous sequence of shapes. This sequence was composed of a concatenation of three-shape items (just as the experiments reviewed above used concatenations of three-syllable non-sense words). Following such a familiarization, participants were as good at discriminating high-TP items from low-TP items when the items were played forward (that is, in the temporal order in which they had been seen during familiarization) as when they were played backwards. If a preference for high-TP items implied that these items have been extracted and memorized, one would have to conclude that participants have extracted also the backwards items although they had never seen them. It thus seems that a preference for high-TP items does not imply that these items have been memorized.

There are also other reasons to doubt that TPs may play an important role in word-segmentation. One reason is that computational studies using TPs (or related statistics) for segmenting realistic corpora of child-directed speech have encountered mixed success at best (e.g., Swingley, 2005; Yang, 2004). At minimum, TPs thus have to be complemented with other cues. This seems highly plausible, given that one would certainly not expect a single cue to solve a highly complex problem such as speech-segmentation.

While the poor performance of word-segmentation mechanisms based on TPs can be improved by the inclusion of other cues, there is a second, more fundamental, reason for doubting that TPs play an important role in word-segmentation. This reason is related to the kinds of

[1] Other speech segmentation models track "chunks" that occur in the input (e.g., Batchelder, 2002; Perruchet & Vinter, 1998). However, as these models have received less experimental attention and make the same predictions for the purposes of the current experiments as the transitional probability-based models, we will not discuss them further.