Journal of Memory and Language 66 (2012) 807–818

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

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

Beyond transitional probability computations: Extracting word-like units when only statistical information is available

Pierre Perruchet *, Bénédicte Poulin-Charronnat

CNRS-UMR 5022, University of Bourgogne, Dijon, France

A R T I C L E I N F O

Article history:
Received 15 April 2011
revision received 17 February 2012
Available online 21 March 2012

Keywords:
Statistical learning
Artificial language
Word segmentation
Chunking
Modeling

A B S T R A C T

Endress and Mehler (2009) reported that when adult subjects are exposed to an unsegmented artificial language composed from trisyllabic words such as AB*X*, *Y*BC, and A*Z*C, they are unable to distinguish between these words and what they coined as the "phantom-word" ABC in a subsequent test. This suggests that statistical learning generates knowledge about the transitional probabilities (TPs) within each pair of syllables (AB, BC, and A···C), which are common to words and phantom-words, but, crucially, does not lead to the extraction of genuine word-like units. This conclusion is definitely inconsistent with chunk-based models of word segmentation, as confirmed by simulations run with the MDLChunker (Robinet, Lemaire, & Gordon, 2011) and PARSER (Perruchet & Vinter, 1998), which successfully discover the words without computing TPs. Null results, however, can be due to multiple causes, and notably, in the case of Endress and Mehler, to the reduced level of intelligibility of their synthesized speech flow. In three experiments, we observed positive results in conditions similar to Endress and Mehler after only 5 min of exposure to the language, hence providing strong evidence that statistical information is sufficient to extract word-like units.

Introduction

Seminal studies have shown that, after hearing an artificial language in which invented words have been concatenated without any phonological or prosodic markers, infants (Saffran, Aslin, & Newport, 1996), children (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), and adults (Saffran, Newport, & Aslin, 1996) become more familiar with the invented words of the language than with the part-words straddling word boundaries. The statistical structure being the only cue made available to the learners, this achievement attests to the fact that listeners are able to exploit the statistical information available in the input, and more precisely, the prevalent view is that this form of learning proceeds through the computation of transitional probabilities (hereafter: TPs; Aslin, Saffran, & Newport, 1998). Participants would exploit the fact that TPs between word internal syllables are stronger than TPs between syllables spanning word boundaries.

The role of statistical information in word segmentation

Since this earlier demonstration, the role of statistics in word extraction has been keenly challenged. A part of the debate stems from the a priori argument that statistical information would be too impoverished to be useful in word learning. For instance, Yang (2004) reported that using TPs leads to a far from optimal segmentation of a child-directed corpus of language: Precision was 41.6%, meaning that more than half of the extracted units were not words, and completeness was 23.3%, meaning that almost 80% of the actual words were not extracted.

Other authors have argued that many other sources of information are available to infants. The role of phonological and prosodic features, such as lexical stress placement,

* Corresponding author. Address: Université de Bourgogne, LEAD/CNRS, Pole AAFE, Esplanade Erasme, 21000 Dijon, France.
E-mail address: pierre.perruchet@u-bourgogne.fr (P. Perruchet).

0749-596X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2012.02.010

on word discovery has been well documented (e.g., Creel, Tanenhaus, & Aslin, 2006; Curtin, Mintz, & Christiansen, 2005; Thiessen & Saffran, 2007). The question of how statistical and phonological or prosodic cues combine has been investigated in experimental studies in which these cues are either consistent or inconsistent with the word-like units of a continuous speech flow (e.g., Creel et al., 2006; Onnis, Monaghan, Chater, & Richmond, 2005; Shukla, Nespor, & Mehler, 2007; Tyler, Perruchet, & Cutler, 2006). These studies have shown that performance in a word-segmentation test improved with consistent cues and strongly decreased (and potentially dropped at chance level) with inconsistent cues. Other studies have explored how these cues compete as function of age. Although Thiessen and Saffran (2003) reported a prevalence of statistics over prosody in 6-month-old infants, Johnson and Jusczyk (2001) reported that prosodic factors override statistics in 8-month-old infants.

Natural language acquisition also relies on the exploitation of known words to discover new words. To borrow an example given by Dahan and Brent (1999): "If *look* is recognized as a familiar unit in the utterance *Lookhere!* then *look* will tend to be segmented out and the remaining contiguous stretch, *here*, will be inferred as a new unit" (p. 165). Dahan and Brent (1999) and Perruchet and Tillmann (2010) provided experimental evidence of this phenomenon in adults, and Bortfeld, Morgan, Golinkoff, and Rathbun (2005) demonstrated the same capacity in 6-month-old infants. It has been suggested that such lexically-driven segmentation could progressively supersede cues during language development (e.g., Mattys, White, & Melhorn, 2005, p. 493).

Although these studies admittedly reduce the relative importance of statistical learning, they did not challenge the ability of statistical learning processes to trigger word-unit extraction when only statistical information is available. In a recent paper, Endress and Mehler (2009) went far beyond this earlier literature. They acknowledged the capacity of learners to compute TPs, but, "surprisingly", they wrote, "there is no evidence that TP-based computations lead to the extraction of word-candidates." The available experimental evidence, they claimed, "does not imply that the items with stronger TPs are represented as actual word-like units, or even that they have been extracted." (p. 352).

Endress and Mehler's (2009) results

Endress and Mehler (2009) based their conclusion on a set of ingeniously designed experiments in which participants were familiarized with a continuous language containing trisyllabic words, as in the studies cited above, but the words were generated from what the authors coined as a "phantom-word", which was never presented in the language. If the phantom-word is designated as ABC (with each letter standing for a syllable), the heard words were AB*X*, *Y*BC, and A*Z*C (with *X*, *Y*, and *Z* standing for invariant syllables). For instance, participants heard *tazepi*, *mizeRu*, and *tanoRu*, which were all derived from the (unheard) phantom-word *tazeRu*. In this way, the phantom-words had exactly the same TPs between their constituent syllables (i.e., AB, BC, and A···C) than the trisyllabic words composing the language. The reasoning was straightforward: If subjects have learned a word-like unit, that is some acoustical word candidates that could be later associated as a whole to a meaning, they should select words over phantom-words when both are played in a subsequent forced-choice test. However, if they only learned pairwise relations, they should be unable to distinguish between the actual words and the phantom-words.

The results indicate that participants failed to distinguish between words and phantom-words. Chance performance was observed in several experiments in which the number of words and the length of the familiarization phase (from 5 to 40 min) were varied. To quote the authors: "Even when collapsing all 161 participants who took part in the different experiments, no preference for words to phantom-words emerged (M = 51.2%, SD = 19.4%), $t(160) = 0.8$, $p = .438$)" (p. 358). In subsequent experiments, the authors made the word structure perceptually salient, either by introducing 25-ms silent pauses between words or by lengthening the last syllable of each word during the familiarization phase. Subjects now chose words over phantom-words in subsequent forced-choice tests. According to the authors, these findings demonstrate that "despite the general agreement that TP-based computations are crucial for word-learning, other cues seem to be required for actually extracting word-like units." (p. 359). In their view, extracting word-like units requires the construction of positional memories, which would be possible only when prosodic markers of word boundaries are provided in the input.

Theoretical implications of Endress and Mehler's (2009) conclusion

Insofar as Endress and Mehler's conclusion is taken for granted, it should elicit major changes in the current conceptions about the role of statistical learning in word segmentation and language acquisition. We focus below on their theoretical implications with regard to learning models. In the brief outline of the statistical approach above, we have assumed that chunks are inferred from the discovery of the boundaries, which are defined as the points where the predictability of the next element is the lowest. Because the primary aim of computations is to insert word boundaries within a continuous sequence, this view, which is currently prevalent in the literature, is sometimes coined as the *bracketing* approach[1] (Goodsitt, Morgan, & Kuhl, 1993; Swingley, 2005). The consequences of Endress and Mehler's results for a bracketing approach are relatively limited. Indeed, Endress and Mehler do not put into question the fact that learners compute TPs, which are at the core of the bracketing approach. They only challenge the additional postulate according to which TP computations directly trigger word-like unit extraction.

[1] The term "statistical learning" sometimes refers to this specific approach, and is taken as equivalent to "computations of transitional probabilities". Hereafter, "statistical learning" is used as a theoretically neutral label designating any form of exploitation of the statistical structure of the input. The chunk-based models described in this paper are construed as models of statistical learning, as those relying on the computation of TPs.