Neuropsychologia 98 (2017) 56–67

Contents lists available at ScienceDirect

Neuropsychologia

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

Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning

Clément François[a,b,c,*], Toni Cunillera[b], Enara Garcia[a,b], Matti Laine[d], Antoni Rodriguez-Fornells[a,b,e,*]

[a] Cognition and Brain Plasticity Group [Bellvitge Biomedical Research Institute-] IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain
[b] Department of Cognition, Development and Educational Science, University of Barcelona, Barcelona, Spain
[c] Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Spain
[d] Department of Psychology, Abo Akademi University, Turku, Finland
[e] Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain

ARTICLE INFO

Keywords:
Speech segmentation
Word referent association
Audiovisual
Statistical learning
Event-related brain Potentials
FN400

ABSTRACT

Learning a new language requires the identification of word units from continuous speech (the speech segmentation problem) and mapping them onto conceptual representation (the word to world mapping problem). Recent behavioral studies have revealed that the statistical properties found within and across modalities can serve as cues for both processes. However, segmentation and mapping have been largely studied separately, and thus it remains unclear whether both processes can be accomplished at the same time and if they share common neurophysiological features. To address this question, we recorded EEG of 20 adult participants during both an audio alone speech segmentation task and an audiovisual word-to-picture association task. The participants were tested for both the implicit detection of online mismatches (structural auditory and visual semantic violations) as well as for the explicit recognition of words and word-to-picture associations. The ERP results from the learning phase revealed a delayed learning-related fronto-central negativity (FN400) in the audiovisual condition compared to the audio alone condition. Interestingly, while online structural auditory violations elicited clear MMN/N200 components in the audio alone condition, visual-semantic violations induced meaning-related N400 modulations in the audiovisual condition. The present results support the idea that speech segmentation and meaning mapping can take place in parallel and act in synergy to enhance novel word learning.

1. Introduction

Learning a new language requires the complex task of isolating new auditory word-forms and associating them onto meanings. Several learning mechanisms have been proposed to explain how adults and infants can solve the word-to-world mapping problem (Kuhl, 2004; Davis and Gaskell, 2009; Rodriguez-Fornells et al., 2009). In particular, statistical learning (SL) or the ability to track regularities or patterns of various sorts in the input seems to be one of the core learning mechanisms for language acquisition, allowing human infants and adults to decipher the units contained in continuous speech streams even after a brief exposure (Saffran et al., 1996). Probably based on the idea that infants need to segment the auditory input first to then attribute meaning to the isolated words, the segmentation and mapping processes have been originally studied in a sequential manner. For instance, the first behavioral studies on segmentation and mapping were conducted in both infant and adult participants who were asked to perform a classic segmentation task followed by a consecutive word-picture mapping task (Mirman et al., 2008; Graf et al., 2007; Hay et al., 2011). Results of these studies indicated that the statistical learning mechanism creates possible word candidates ready to be mapped onto meaningful representations. These memory traces stemming from statistical learning may thus constitute a proto-vocabulary that is gradually created before being mapped to the corresponding conceptual units (Fernandes et al., 2009; Rodriguez-Fornells et al., 2009). This mapping could be accomplished by associative or statistical learning mechanisms that would bind conceptual representations onto these proto-vocabulary traces.

A recent article presents an alternative view on the original idea that segmentation and mapping processes operate in a sequential manner (Räsänen and Rasilo, 2015). These authors propose a computational model of joint word segmentation and meaning mapping based

* Corresponding authors at: Department of Basic Psychology, University of Barcelona, Campus de Bellvitge - Pavelló de Govern, 08908 L'Hospitalet de Llobregat, Barcelona, Spain.
E-mail addresses: fclement24@hotmail.com (C. François), antoni.rodriguez@icrea.cat (A. Rodriguez-Fornells).

http://dx.doi.org/10.1016/j.neuropsychologia.2016.10.006
Received 29 January 2016; Received in revised form 3 October 2016; Accepted 8 October 2016
Available online 11 October 2016
0028-3932/

on empirical and simulated data in which a shared domain-general statistical learning mechanism may act both within and across modalities. Importantly, these authors propose that the learning process might be facilitated by the simultaneous appearance of the word and its visual referent. Indeed, new words are usually heard in different contexts and in some cases with the simultaneous appearance of the external referent associated to the new auditory word-form. Binding new-words and possible referents across different contexts may rely on a bootstrapping mechanism that helps to fill the language-learning gap. The statistical properties found in the speech signal as well as the statistical consistency between speech and the world (external context and referents) might be therefore crucial to infer the possible meaning of a new word and also to help the speech segmentation process.

In line with this idea, Cunillera and colleagues (2010a) conducted a behavioral study using an audiovisual learning paradigm in which a word-segmentation task was coupled with a word-picture association task with varying degrees of association consistency and meaningfulness. In contrast to the sequential proposal that meaning mapping would occur after speech segmentation (Graf et al., 2007; Hay et al., 2011), the results of this study clearly showed that segmentation and word-referent mapping could be accomplished after a very short exposure, most likely in a simultaneous fashion. Word learning was more effective when visual referents were meaningful objects and speech segmentation performance increased with the presence of systematic word-picture associations. The benefit of visual cues on auditory segmentation was further confirmed with different types of multisensory cues shown to boost speech segmentation (Thiessen, 2010; Glicksohn and Cohen, 2013). These data are in agreement with the observation that the prosodic structure of child-directed speech offers the possibility to solve the speech segmentation and word-to-world mapping problem in parallel (Yurovsky et al., 2012). Thus, several converging evidences and theoretical proposals concur by pointing out that both processes, speech segmentation and meaning mapping, could be functionally active in parallel during learning. However, all these studies based their interpretation on the analysis of the behavioral data collected after the learning has taken place.

In the present study, we went one step further by investigating the neurophysiological mechanisms involved when the two processes of segmenting new words and binding conceptual representations onto these newly isolated words are taking place at the same time. One crucial component of the Event-Related Potentials underlying semantic-conceptual processing is the N400 component (Kutas and Hillyard, 1980). Classical views of the central-parietal N400 associate its amplitude modulations to lexical and semantic retrieval processes (Kutas and Federmeier, 2000). Nonetheless, recent ERP studies have provided converging evidence for a fronto-central N400-like component (FN400) in artificial language learning tasks with increasing amplitude during the initial stages of extracting new words (Cunillera et al., 2009; De Diego-Balaguer et al., 2007; François et al., 2014). Interestingly, the topographical distribution of the learning-related N400 component suggests the involvement of specific neurophysiological mechanisms indexing speech segmentation that differ from the lexical-semantic retrieval process (see also Dittinger et al. (2016)). Similar frontal N400 components were observed in 14-month-old infants performing a word-picture mapping task (Friedrich and Friederici, 2008; see also for similar ERP results in 17–21 month-old infants, Mills et al. (2005)). Frontally distributed N400 were also observed in language-learning studies using semantic priming with newly learned words (Mestres-Missé et al., 2007). Moreover, a similar FN400 has been recently associated to conceptual/semantic priming processes in several studies of memory recognition (see for a discussion, see Voss and Federmeier (2011), Voss et al. (2010)). Therefore the FN400 seems to be a good candidate for studying speech segmentation and word-picture association processes notably when occurring in parallel. Here, we used this FN400 (i) to explore the neurophysiological mechanisms of speech segmentation and word-picture mapping when occurring in parallel during an initial learning phase and, (ii) to collect implicit brain responses of incorrect visual word-picture associations during an implicit test phase.

Besides the FN400, we took advantage of two other ERP components of interest, namely the Mismatch Negativity (MMN) and the N200. The MMN is a sensitive measure of pre-attentive, automatic and implicit auditory change detection, which may reflect the formation of an echoic memory trace within the auditory cortex (Näätänen et al., 2005). The MMN is elicited by rare stimuli presented in a sequence of repeated standard stimuli (Näätänen et al., 2005) and is observed for simple (Näätänen et al., 1978, 2005; Deguchi et al., 2010; Chobert et al., 2012) as well as for more complex auditory patterns of speech and non-speech stimuli (Boh et al., 2010; Herholz et al., 2009; Wang et al., 2012). The N200 component has been observed in artificial grammar learning studies with ungrammatical sequences eliciting larger N200 than grammatical sequences (Carrion and Bly, 2007; Selchenkova et al., 2014). The N200 has been also related to working memory processes underlying template mismatch mechanisms (Sams, Alho and Näätänen, 1983) and may thus index the acquisition of implicit knowledge (Selchenkova et al., 2014). Therefore, after a learning phase we exposed participants to test streams composed of standard tri-syllabic words and infrequent structurally illegal tri-syllabic words to collect implicit measures of structural auditory change detection (for a similar procedure, see De Diego-Balaguer et al. (2007)).

In order to explore the neurophysiological mechanisms of speech segmentation and binding new-words onto conceptual representations, we exposed participants to a similar multimodal statistical learning paradigm as used by Cunillera et al. (2010a). Auditory streams composed of statistically concatenated new artificial tri-syllabic words were presented with (Audiovisual condition) or without (Audio alone condition) consistently associated visual information while we recorded EEG (see Fig. 1 for the experimental design). For both the audio alone and audiovisual (AV) conditions, participants were also exposed to a baseline in which all the syllables appeared randomly, thus providing no statistical cue to word boundaries. Importantly, in the random streams of the AV condition both the syllables and the visual referents were randomly presented. After the learning phases, we collected brain activity during an implicit test in which structural (Audio alone condition) or visual binding mismatches (AV condition) were pseudo-randomly inserted in the streams. Behavioral measures of word recognition and word-picture associations were finally collected. For the learning phases of both conditions, we compared ERPs elicited by structured streams to those elicited by random ones. We hypothesized that the AV streams should provide facilitatory cues in the segmentation process by allowing participants to bind possible meaning representations onto newly segmented words (see Fig. 1). The facilitation provided by this additional mapping process should be accompanied by enhanced word recognition performance during the explicit behavioral test. At the electrophysiological level, we expected to observe a similar FN400 in the two conditions if simultaneous segmentation and mapping involve similar neurophysiological mechanisms. On the other hand, if the simultaneous tracking of words and their referents involve an increased cognitive load, we expected to observe larger and/or later FN400 in the AV than in the audio alone condition. In order to test for these hypotheses, we performed a complementary analysis on the FN400 effects observed in the two conditions to provide direct evidence for the involvement of different neurophysiological mechanisms in the two processes. During the implicit test phase of the audio alone condition, we expected auditory structural mismatches to elicit MMN/N2 components (Carrion and Bly, 2007; Selchenkova et al., 2014). On the contrary, we expected the visual binding mismatches to elicit modulation of the lexical-semantic N400 potential during the implicit test phase of the AV condition (Ganis et al., 1996).