Cognition 181 (2018) 135–140

Contents lists available at ScienceDirect

Cognition

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

Brief article

Cortical tracking of constituent structure in language acquisition

Heidi Getz[a,*,1], Nai Ding[b,1], Elissa L. Newport[a], David Poeppel[c,d]

[a] Center for Brain Plasticity and Recovery, Department of Neurology, Georgetown University, Building D Suite 154, 4000 Reservoir Road NW, Washington, DC 20057, USA
[b] Zhejiang University, Hangzhou, People's Republic of China
[c] New York University, Department of Psychology, 6 Washington Place, New York, NY 10003, USA
[d] Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

ARTICLE INFO

Keywords:
Statistical learning
Syntax
Hierarchical structure
Entrainment
MEG

ABSTRACT

Linguistic units are organized at multiple levels: words combine to form phrases, which combine to form sentences. Ding, Melloni, Zhang, Tian, and Poeppel (2016) discovered that the brain tracks units at each level of hierarchical structure simultaneously. Such tracking requires knowledge of how words and phrases are structurally related. Here we asked how neural tracking emerges as knowledge of phrase structure is acquired. We recorded electrophysiological (MEG) data while adults listened to a miniature language with distributional cues to phrase structure or to a control language which lacked the crucial distributional cues. Neural tracking of phrases developed rapidly, only in the condition in which participants formed mental representations of phrase structure as measured behaviorally. These results illuminate the mechanisms through which abstract mental representations are acquired and processed by the brain.

1. Introduction

Linguistic units are organized at multiple levels, producing layers of structure: word combinations form phrases, which combine to form sentences. Continuous speech lacks definitive physical cues to the boundaries between these units (Lehiste, 1970; Morgan & Demuth, 1996). Nevertheless, recent experiments reveal that the brain tracks the presentation of linguistic units in real time, "entraining" to multiple levels of hierarchical structure simultaneously.

Ding, Melloni, Zhang, Tian, and Poeppel (2016) recorded magnetoencephalography (MEG) data while native speakers of English or Mandarin Chinese listened to sequences of words, phrases, or sentences in each language. The units at each level of organization occurred periodically, at a specific frequency. The neural response to units at each hierarchical level was extracted by calculating the MEG power spectrum at each frequency. Results revealed concurrent tracking (time-locked neural activity) of phrases and sentences in the native language, but not in a foreign language. The authors emphasized that neural tracking reflects knowledge of an abstract mental grammar rather than low-level statistical information. However, this grammar was acquired through a learning process. At some earlier stage there must have been a transition where learners began to represent serially ordered material hierarchically. When during this transition does the brain begin to track hierarchical structure in real time?

Learners can organize continuous speech into smaller units through statistical learning. For example, learners use transition probabilities between syllables to identify word boundaries (Saffran, Aslin, & Newport, 1996). However, knowledge of these boundaries does not necessarily lead to neural tracking of the corresponding units. Buiatti, Peña, and Dehaene-Lambertz (2009) report neural tracking of newly segmented words only when boundaries were marked with subliminal (25 ms) pauses, and not in conditions without pauses, even though participants discriminated words from part-words in both conditions. Subliminal pauses also marked word boundaries in a study by Kabdebon, Pena, Buiatti, and Dehaene-Lambertz (2015) on neural tracking in infants. One adult statistical learning study (Batterink & Paller, 2017) reported tracking without perceptual cues, but that study used simple and repetitive materials designed by Saffran et al. (1996) for 8-month-old infants. (The entire 12-minute exposure session consisted of four repeating trisyllabic words, which infants learn after two minutes.) Neural tracking of this highly repetitive speech stream seems different from the type of response reported by Ding, Melloni, Zhang, et al. (2016), which varied with the abstract hierarchical structure of the materials. Indeed, Batterink and Paller (2017) report that the same

* Corresponding author at: Center for Brain Plasticity and Recovery, Department of Neurology, Building D, Suite 145, 4000 Reservoir Road NW, Washington, DC 20057, USA.

E-mail address: hrg2@georgetown.edu (H. Getz).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cognition.2018.08.019
Received 26 September 2017; Received in revised form 28 August 2018; Accepted 29 August 2018
0010-0277/

learners also tracked sequences of trisyllables that did not form words, and interpret tracking in their study as a perceptual phenomenon.

Thus, neural tracking of abstract linguistic structure in the absence of perceptual cues has been convincingly demonstrated only in native speakers of natural languages (Ding, Melloni, Zhang, et al., 2016). Such abstract linguistic representations must have been acquired from something more concrete. When during this transition does neural tracking begin? Does neural tracking emerge only after extensive experience with a natural language, which would suggest that it depends on well-established linguistic representations? Or does neural tracking emerge as soon as learners can identify linguistic units in the speech stream? Our study addresses this question using a miniature language paradigm that leads adult learners to form abstract representations of phrases.

1.1. The present study

We measured neural activity using MEG using the technique developed by Ding, Melloni, Zhang, et al. (2016) while adults learned a miniature language designed by Thompson and Newport (2007). In the target condition, sentences contained a number of language-like distributional cues to phrase structure. In a control condition, sentences lacked such cues. In Thompson and Newport's study, only learners in the target condition formed phrase-structure representations of the sentences in the miniature language. Here we asked whether learners in this condition would neurally track the acquired phrase structure, and if so, when this response would emerge.

2. Methods

2.1. Design

2.1.1. Transitional probabilities cueing phrase structure

In natural languages, syntax operates on phrasal constituents, with the words within phrases acting as a unit (e.g., words within prepositional phrases can be omitted or moved together: *The box (on the counter) is red. The box is sitting (on the counter)*). Over a corpus, the consistent syntactic behavior of words within phrases creates a specific pattern of transitional probabilities between word categories.[2] Categories that form a phrase (such as prepositions, articles, and nouns: *on the counter*) tend to have high transitional probabilities; categories that span a phrase boundary (such as nouns and verbs: *counter is*) have lower transitional probabilities. Thompson and Newport (2007) showed that learners use this contrast between highly probable and less probable sequences of categories to identify which categories form phrases in a miniature language. No acoustic, prosodic, or semantic cues were necessary.

2.1.2. Miniature language structure

The language was identical to that in Experiment 4 of Thompson and Newport (2007). The basic sentence structure is ABCDEF. Each letter represents a grammatical category (such as Noun) with 2 or 4 monosyllabic lexical items (e.g., "hox", "lev"). We formed complex sentences by applying syntactic patterns found in natural languages (omission, repetition, movement, or any combination of these) to pairs of words in the basic structure (Table 1). To create the phrase-structure language, all of these transformations were applied to consistent pairs of words (AB, CD, and EF). This produced the critical pattern of high transitional probability within phrases and dips in transitional probability at phrase boundaries. Within phrases in our miniature language, transitional probabilities between categories are perfect: every A word is always followed by a B word, every C word is followed by a D word, and so on.[3] Between phrases, transition probabilities are lower, since, for example, B words can be followed by C words, E words, or A words. A control language without phrase structure was created by applying the same transformations to any adjacent pair of words. Because no word categories were consistently grouped together (e.g., A could be followed by B, D, or F), there were no peaks and dips in transitional probabilities and no cues from these statistics to phrase structure.

The exposure set for each language consisted of 39 sentences, 5% with the basic ABCDEF structure and 95% with a transformed structure. The sentences in each exposure set were randomly ordered, and this ordered set was looped 18 times.

2.1.3. Phrase test

A two-alternative forced-choice test assessed whether participants had learned to group words into phrases. On each of 18 trials, participants heard two sequences of words. Both choices were legal two-word sequences and occurred during exposure for both languages. In the phrase-structure language, one sequence formed a phrase (AB) while the other spanned a phrase boundary (BC). The phrase occurred first or second equally often. Neither sequence formed a phrase in the control language because that language did not have phrases. Participants were asked to choose the pair that formed a better group in the language. The phrasal sequence was considered "correct" for scoring purposes.[4]

2.2. Materials

We synthesized individual monosyllabic words using the Alex voice in MacInTalk, without prosody, and edited them to control acoustic features.[5] Words were then concatenated to create 39-sentence (132 s) exposure sets. This process was blind to sentence structure and did not produce acoustic cues at phrase boundaries. Sentences were separated by 784 ms silence (the duration of a phrase). Each exposure set was looped 18 times, for a total of 40 min of exposure.

Because the languages shared a vocabulary and because the process for creating sentences did not physically alter these words, the two exposure sets had the same basic acoustic properties. Therefore, any differences in neural tracking across conditions cannot be attributed to physical differences in the materials and must instead reflect higher-level knowledge of the language's structure–information that is not encoded in the physical speech signal.

2.3. Procedure

Thirty-two adults at New York University (right handed, 15 male, mean age 23, SD = 4) listened to the phrase-structure language (Group PS: n = 16) or the control language (Group C: n = 16) while we

[2] Transitional probability is a statistic that measures the predictiveness among adjacent elements. The forward transitional probability of successive elements XY is defined as the probability of XY divided by the probability of X.

[3] Since each category contains 2 or 4 words, word-level transition probabilities will differ from category-level transition probabilities. Peaks and dips in transition probability are clearest when defined across categories, as in natural languages.

[4] One of the items was 'correct' in the phrase structure condition only if learners did acquire phrase structure from the transition probability cues. However, as noted, both choices were legal sequences in both languages, and in the control condition there were no cues to phrase structure and therefore no reason for participants to prefer one choice over another. Performance on this test in the control condition thus served as a baseline for the phrase-structure condition, to ensure that participants did not perform correctly for extraneous reasons.

[5] Words were synthesized to durations of 380–390 ms. Each word was subsequently edited to 380 ms (by trimming or adding silence) and normalized for loudness and power. Offsets were smoothed with a 25 ms cosine window. Words were assigned to categories such that there were no phrasal rate acoustic fluctuations over the corpus as a whole.