Cortical tracking of constituent structure in language acquisition

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

[a]Center for Brain Plasticity and Recovery, Department of Neurology, Georgetown University, Building D Suite 145, 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

[#] These authors contributed equally to this work.

Abstract

Linguistic units are organized at multiple levels: words combine to form phrases, which combine to form sentences. Ding et al. (2016a) 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.

Keywords

statistical learning; syntax; hierarchical structure; entrainment; MEG

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 et al. (2016a) 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 et al. (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 et al. (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 et al. (2016a), which varied with the abstract hierarchical structure of the materials. Indeed, Batterink and Paller (2017) report that the same 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 et al. 2016a). 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 et al. (2016a) 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.[1] 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 phrasestructure 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.[2] 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.

---

[1]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.
[2]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.

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 phrasestructure 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.[1]

2.2 Materials

We synthesized individual monosyllabic words using the Alex voice in MacInTalk, without prosody, and edited them to control acoustic features.[2]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 minutes 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 recorded MEG. Subjects were invited to take a brief break halfway through exposure. To ensure attention, participants were asked to press a button after every 4-word sentence. Participants fixated on a target to reduce eye movement artifacts in the MEG signal. After the 40minute exposure session, participants completed the Phrase Test described in 2.1.3.

2.4 MEG analysis

MEG recordings were made with a 157-sensor system (KIT). Signals were sampled at 1 kHz (200 Hz low-pass filter, 60 Hz notch). The neural signals were bandpass filtered between 0.7 Hz and 6 Hz using a FIR filter and resampled to 200 Hz offline before analysis. Environmental noise was removed using time-shift principal component analysis (de Cheveigne & Simon, 2007). The neural recordings during exposure were segmented into 64 38s epochs. For each MEG channel, values deviating from the mean by more than 9 SD

were considered saturated. To extract response components phase-locked to the stimulus, we decomposed the data using denoising source separation (DSS); the bias response was chosen as the response averaged over all 38s epochs (de Cheveigne & Simon, 2008). The first DSS component, which has the most reliable response waveform across epochs, was used for further analysis. To eliminate transient neural responses evoked by sentence boundaries, the neural recordings in a 1520 ms window centered at the sentence onset (covering the 760 ms pause before a sentence and the first two words of each sentence) were removed from the analysis. To calculate the response spectrum with an appropriate frequency resolution, the data were re-epoched into 32 word epochs. MEG signals were transformed into the frequency domain using DFT without any smoothing window, resulting in a frequency resolution of 0.082 Hz. Since DSS components have arbitrary polarities, to average across participants the DSS polarity was aligned using multiset Canonical Correlation Analysis (mCCA) (Zhang & Ding, 2017). The mCCA aligns polarity by maximizing the correlation between the re- sponse waveforms from all pairs of participants.

3 Results

3.1 Neural tracking reflects knowledge of phrase structure

3.1.1 Phrasal knowledge—The Phrase Test measured behaviorally whether subjects acquired the phrase structure of the language. Participants chose between word sequences that, in the phrase-structure language, either formed a phrase (AB) or spanned a phrase boundary (BC). Participants exposed to the phrase structure language (PS) chose the phrase sequence significantly more than the control group (C) (Figure 1; 70% vs. 55%, $t(30) =$ 2.5302, p = .009, one-tailed), while Group C did not differ from chance ($t(30) = 1.4035$, p > .15). These results indicate that Group PS, but not Group C, acquired phrase-structure representations.

3.1.2 Neural tracking—Figure 2 shows the MEG response (power spectrum of the first DSS component, which is a weighted average of data from all MEG sensors) averaged over the 40-minute exposure period. The phrase-rate response at 1.3 Hz was significantly stronger in Group PS than Group C ($t(30) = 4.75$, p < .001, two-sample t-test). Because the words of the two languages were the same—only the sentence structure differed—Group PS's stronger phrase-rate response indicates tracking of the language's phrase structure.[3] The word-rate response was not significantly different between groups ($t(30) = 0.18$, p = 0.85, two-sample t-test).

when listening to the materials, but computing statistics over word classes to determine which pairs consistently co-occur and then identifying those pairs on a test. This is a much more stringent test of whether participants learned *phrases*. The strong performance of Group PS on our test demonstrates that they acquired knowledge of these consistent

---

[3]The weak response present in Group C may reflect a tendency to group words into twoword units. We have repeatedly observed a binary grouping tendency in behavioral experiments; this allowed participants in Group C to surpass chance on the phrase test when 50% of the sentences had a stable word order (ABCDEF; Thompson & Newport, 2007). This tendency may also underlie the weak phrase-rate MEG response in Group C and the very early phrase-rate response in Group PS. However, our behavioral measure requires not just hearing binary groups

groupings, whereas Group C did not. Thus knowledge of these groupings must be driving the much more robust phrasal response in Group PS.

3.1.3 Individual variation—There was more overlap across groups on our behavioral measure than on our neural measure (Figure 3). Nearly all participants in Group PS demonstrated strong neural tracking, regardless of behavioral score, suggesting that the brain response to phrase structure may be less variable than the behavioral one.

3.2 Tracking of phrases increases as knowledge develops

We examined how neural tracking emerged by analyzing the MEG response in 22-second bins (5 to 8 sentences) over the exposure (Figure 4). In Group PS, a robust phrase-rate response built gradually, peaking after approximately three and a half minutes (10 bins). A much smaller phrasal response in Group C plateaued after approximately one minute (3 bins).

To identify when the group difference in phrase-rate response emerged, we compared the phrase-rate response over three-bin (~1 minute) intervals. The groups did not differ significantly during bins 1–3 ($F(1,30) = .43$,$p = .52$) or bins 4–6 ($F(1,30) = .97$, $p = .33$), confirming that group differences were not present initially and instead reflected learning by Group PS that occurred during exposure. A significant between-groups difference emerged in bins 7–9 ($F(1,30) = 4.43$, $p = .04$). In separate unpublished experiments, where we have measured subjects' knowledge behaviorally after varying amounts of exposure to these languages, significant behavioral differences begin to emerge after 8–10 bins of exposure (176–220 seconds, or during the second loop of the 132-second exposure set). These results indicate that, at a group level, brain responses indicating representations of phrase structure appear around the same time as—or perhaps just before— this knowledge begins to manifest behaviorally.

4 Discussion & Conclusions

4.1 Summary

We measured neural tracking of linguistic structure (Ding et al., 2016a) during exposure to a miniature language with distributional cues to phrase structure or to a control language (Thompson & Newport 2007). Tracking of phrases was observed only for learners of the phrase-structure language. Because cues to the phrase boundaries were not present in the acoustic signal, this tracking must reflect abstract knowledge of the sentences' phrasal structure. The results suggest that statistically learned phrases are represented abstractly and constructed online during language processing.

4.2 Time course of neural tracking during learning

Brain responses to phrase structure developed rapidly: within four minutes. In contrast, Buiatti et al. (2009) observed neural tracking of learned words after nine minutes of exposure only when boundaries were marked with subliminal pauses. This suggests that our participants' representations somehow differed from those in that study, such that our participants built representations online while theirs did not. Buiatti et al.'s (2009)

participants could explicitly recall learned words only in the condition containing pauses, suggesting that participants in the nopauses condition learned probabilities over syllable *sequences* (thus distinguished words from part-words) but did not as successfully code the sequence into units. Of course, pauses cannot be required for such coding, since our participants did succeed in learning the higher-level organization of the sentence without pauses (and natural languages do not have pauses between linguistic units). But the distributional cues in our study were different: the words that formed a phrase were not only linearly adjacent, they also appeared, disappeared, and moved together—and perhaps this was critical. Understanding how the brain codes the output of different types of statistical computations is an intriguing area for future research. In general, the rapid development of a phrase-rate response in our study is consistent with other evidence of rapid statistical learning (Saffran et al., 1996; Aslin & Newport 2012, 2014).

In contrast to phrases, tracking of acoustically identifiable words decreased over time in both groups, possibly reflecting loss of attention.[4]

4.3 Local temporal dynamics of neural responses

What aspect of phrase structure is the brain tracking? In an exploratory analysis, we asked whether group differences emerged at specific points during sentence processing. Group differences appeared to be localized temporally to the within-phrase transitions (Figure 5). The same high transition-probability pattern forms a cue to phrase structure in natural languages. This suggests that attention to high-probability transitions—perhaps in anticipation of the completion of a grammatical phrase—could underlie the global phrase-rate response in this experiment as well as in Ding et al. (2016a).

This possibility is intriguing because it holds the potential to refine our understanding of neural tracking. However, caution is warranted.[5] Exploring this pattern in future research could reveal further details about neural tracking of linguistic structure.

4.4 Rules and statistics

Ding et al. (2016a,b) argued that knowledge of abstract phrase structure, not "mere statistics," drives neural tracking. On the other hand, Thompson and Newport (2007) argued that abstract phrase structure is acquired on the basis of statistics such as transitional probabilities. These views may appear contradictory, so we take this opportunity to clarify that they are not. Learners certainly use distributional information to acquire the phrase structure of their language. However, simple statistics such as bigram frequency do not lead to brain responses to phrase structure or to mental representations of that structure (Kam et al., 2008; Thompson & Newport, 2007). Rather, we propose that a correlated set of distributional cues underlies the transition to representing serially ordered material in terms

---

[4]In other studies, increases in tracking of higher-level units alongside decreases in tracking of lower-level units was interpreted as evidence that the higher-level responses "suppressed" the lower-level ones (Buiatti et al., 2009; Batterink & Paller, 2017). However, here the lower-level (word) response decreased whether or not a phrase-level response emerged (the decrease was observed in both groups). One reason our results might differ from those studies is that our exposure period was much longer (40 vs. 9–12 minutes). A second reason might be that tracking in prior studies reflects a shift in perceptual grouping (as suggested by Batterink & Paller, 2017), while in this study tracking reflects the processing of an additional layer of structure.
[5]With a periodic neural response, it is impossible to distinguish a short-latency response to a recent event from a longer-latency response to a temporally more distant event.

of a more abstract grammar. As learners begin to represent their language in terms of phrase structure, we believe that is when they show knowledge of which words form phrases as well as a brain response to those phrases in a stream of speech, as in the present experiment. This transition depends on using complex statistical distributions to acquire such representations; but the crucial product of learning is abstract phrase representations and not merely the statistics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

References

Aslin RN, & Newport EL (2012). Statistical learning: From acquiring specific items to forming general rules. Current Directions in Psychological Science, 21(3), 170–176. [PubMed: 24000273]

Aslin RN, & Newport EL (2014). Distributional Language Learning: Mechanisms and Models of Category Formation. Language Learning, 64(s2), 86–105. [PubMed: 26855443]

Batterink LJ, & Paller KA (2017). Online neural monitoring of statistical learning. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 90, 31–45. [PubMed: 28324696]

Buiatti M, Peña M, & Dehaene-Lambertz G (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. NeuroImage, 44(2), 509–519. [PubMed: 18929668]

de Cheveigne A, Simon JZ. (2008) Denoising based on spatial filtering. J Neurosci Methods, 171: 331–339. [PubMed: 18471892]

de Cheveigne A, Simon JZ. (2007) Denoising based on time-shift PCA. J Neurosci Methods, 165: 297–305. [PubMed: 17624443]

Ding N, Melloni L, Zhang H, Tian X, & Poeppel D (2016a). Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience, 19(1), 158–164. [PubMed: 26642090]

Ding N, Melloni L, Tian X, & Poeppel D (2016b). Rule-based and word-level statisticsbased processing of language: insights from neuroscience. Language, Cognition and Neuroscience, 1–6.

Kabdebon C, Pena M, Buiatti M, & Dehaene-Lambertz G (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. Brain and Language, 148, 25–36. [PubMed: 25865749]

Kam X-NC, Stoyneshka I, Tornyova L, Fodor JD, & Sakas WG (2008). Bigrams and the richness of the stimulus. Cognitive Science, 32(4), 771–787. [PubMed: 21635353]

Lehiste I (1970). Suprasegmentals. M.I.T. Press.

Morgan JL, & Demuth K (Eds.). (1996). Signal to syntax: Bootstrapping from speech to grammar in early acquisition. Mahwah NJ: Lawrence Erlbaum Associates.

Saffran JR, Aslin RN, & Newport EL (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926–1928. [PubMed: 8943209]

Thompson SP, & Newport EL (2007). Statistical Learning of Syntax: The Role of Transitional Probability. Language Learning and Development: The Official Journal of the Society for Language Development, 3(1), 1–42.

Zhang W, Ding N (2017) Time-domain analysis of neural tracking of hierarchical linguistic structures. NeuroImage, 146, 333–340. [PubMed: 27856315]
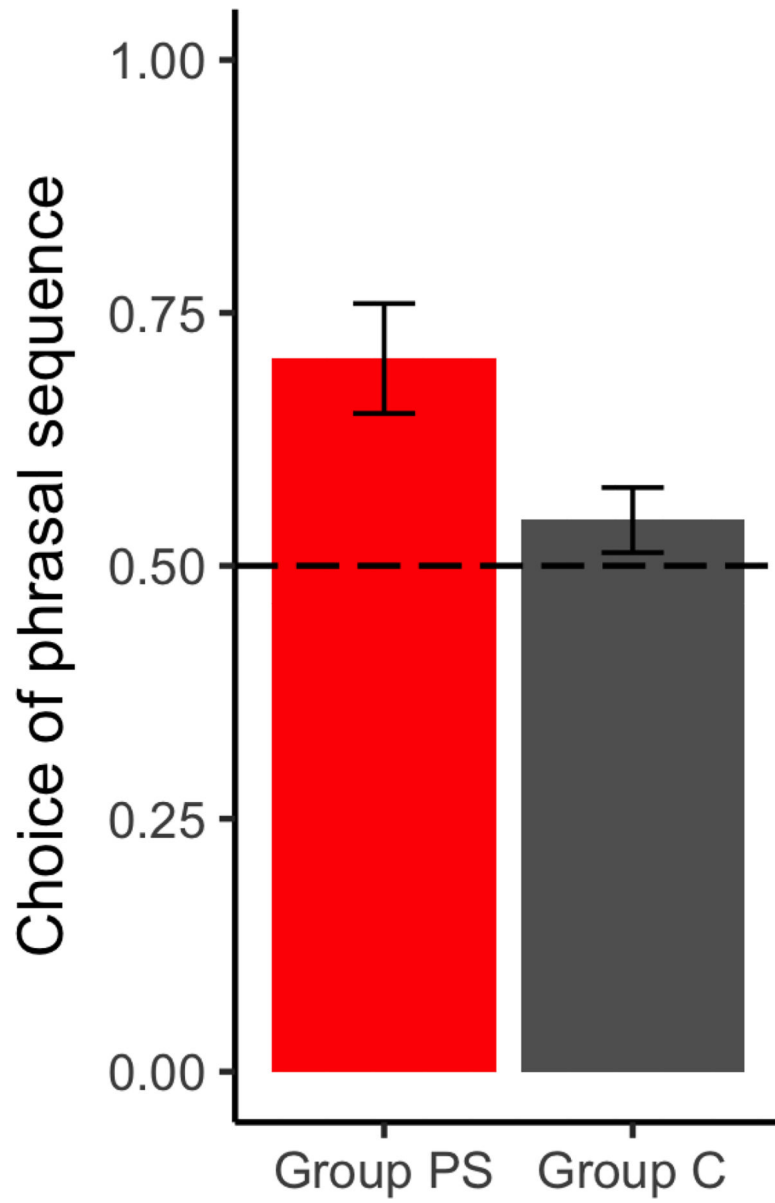
Figure 1.
Performance on phrase test (n=32). Error bars: SEM. Subjects in Group PS preferred phrases to legal non-phrase sequences, indicating phrasal representations.
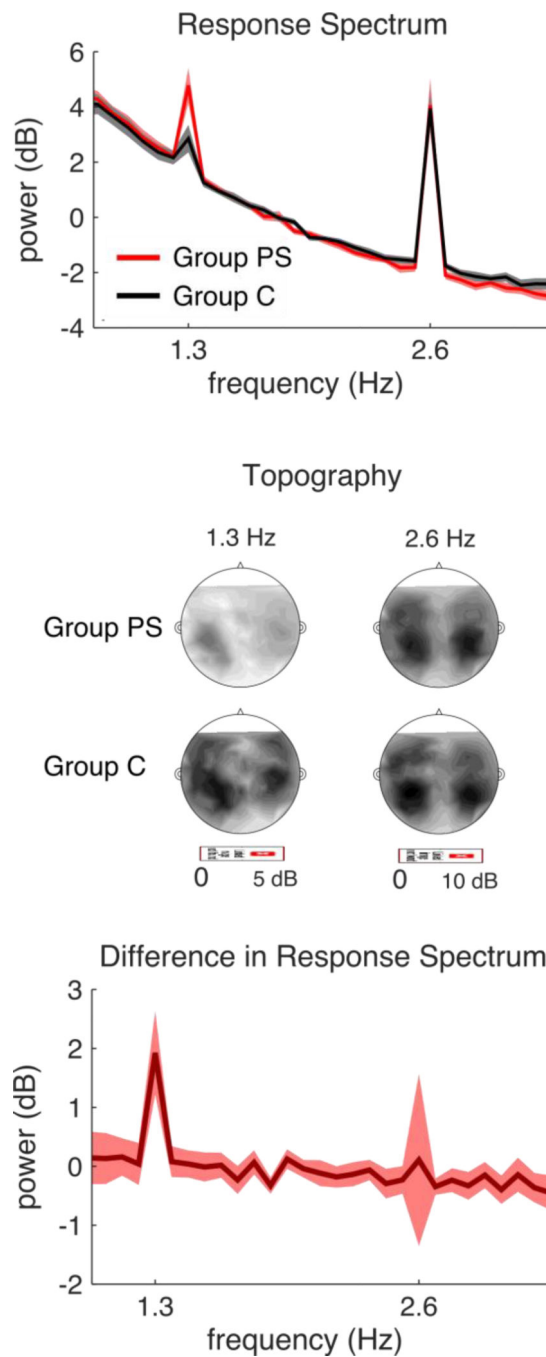
Figure 2.
MEG response and groups differences. The MEG response shown is the first DSS component. The 2.6 Hz peak corresponds to word boundaries, which have physical correlates in the acoustic signal. Both groups track these units. The 1.3 Hz peak corresponds to phrase boundaries, which are not marked physically. Tracking of phrases is strongest in Group PS. Shaded area: 2 SEM over subjects on each side. The response topographies iare shown on the right. Stronger responses are seen for Group PS at 1.3 Hz than Group C. The responses at 2.6 Hz show a similar bilateral pattern for both groups.
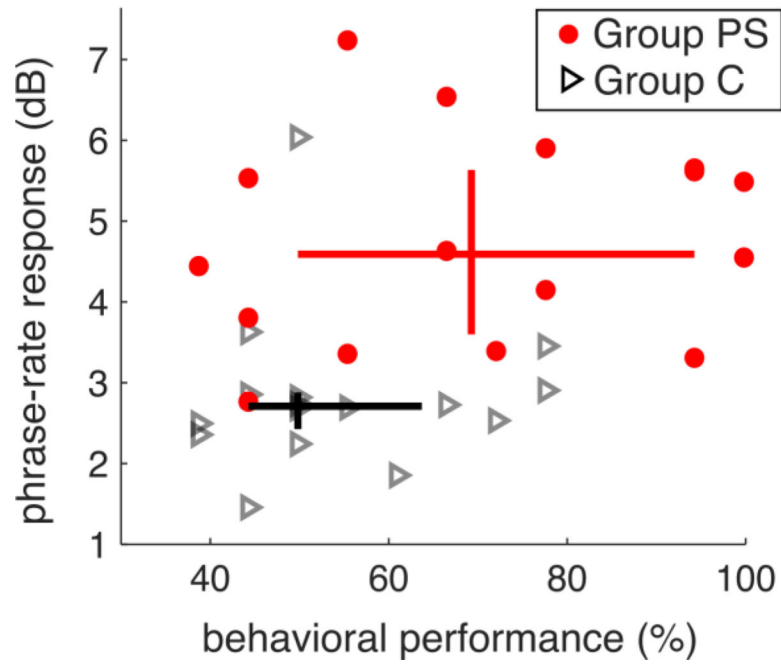
Figure 3.
Relationship between MEG phrasal-rate response and behavioral performance. The Y-axis represents average power at the phrase (1.3 Hz) rate minus average power in a neighboring frequency region (1Hz) in order to correct for differences across subjects. The behavioral performance is the same as shown in Fig. 2. The inter-quartile ranges of behavioral and neural scores are plotted as horizontal and vertical lines respectively. The two horizontal lines overlap, whereas the two vertical lines do not, indicating more pronounced group differences for the neural measure than the behavioral one.
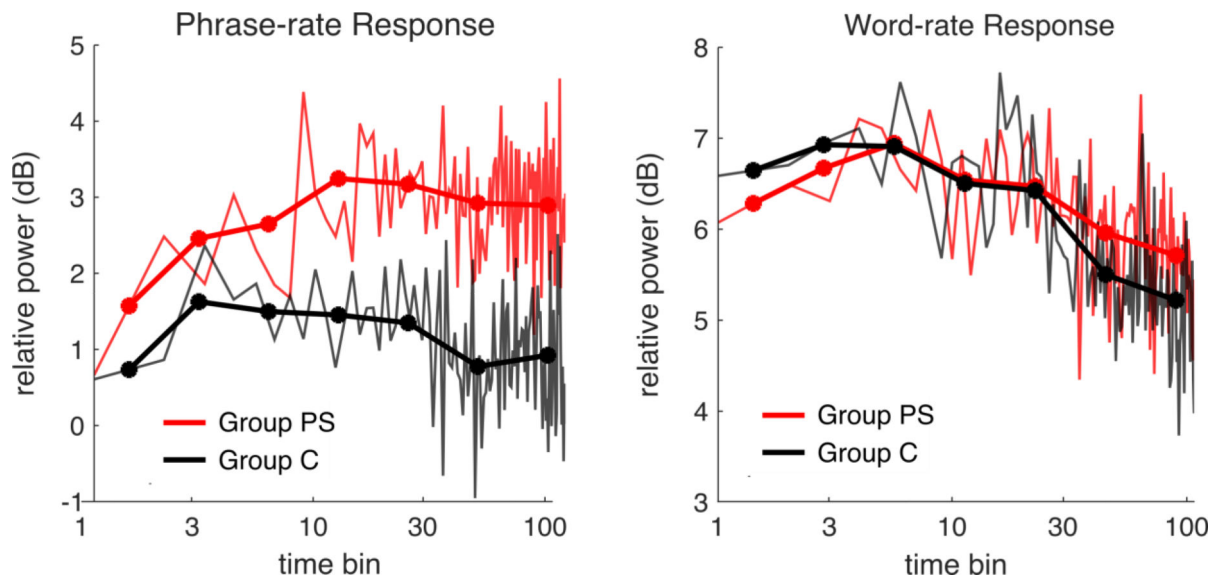
Figure 4.
Building of neural response over time. MEG responses were divided into bins of 22 seconds (graphed on a logarithmic scale). Neighboring bins were averaged into blocks (markers). Left, Phrasal rate; right, word rate. The Y-axis represents average power at the phrase (1.3Hz) rate or word (2.6Hz) rate minus average power in a neighboring frequency region (1Hz), in order to correct for differences across subjects.
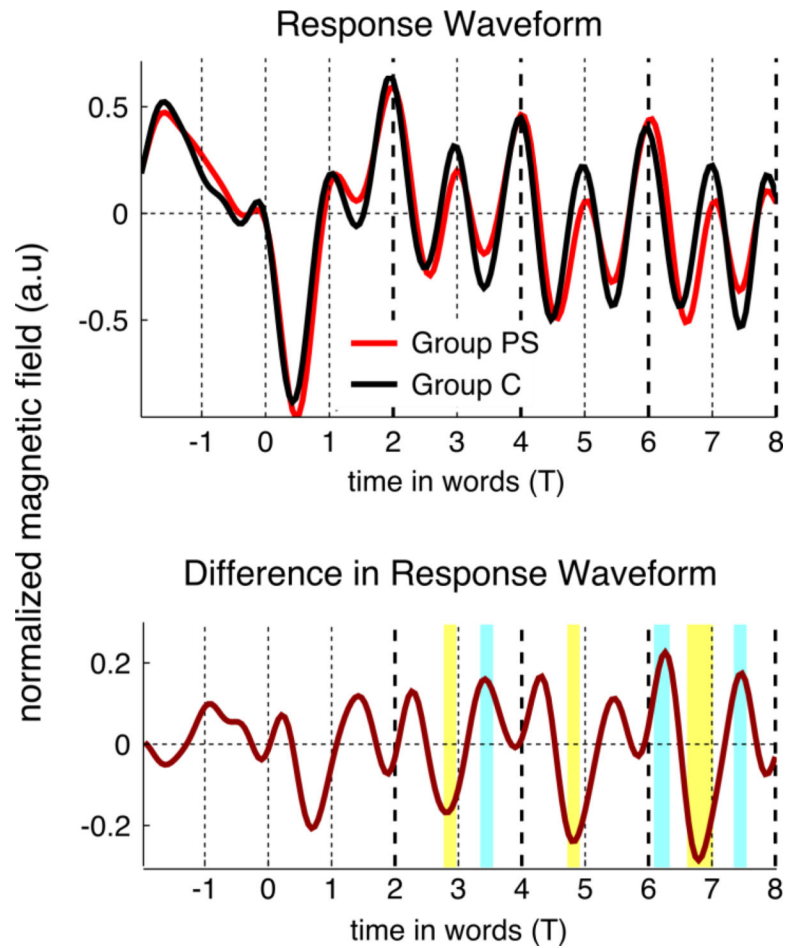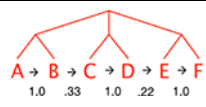
Figure 5.
Temporal dynamics of the MEG response, averaged across the entire exposure period. The offset of each phrase is shown by a thick dashed line, the offset of other words is shown by a thin dashed line. Top: response time course for the two conditions. Bottom: difference between conditions. Significant differences are marked by yellow (Group PS < Group C) and cyan (Group PS > Group C) bars (unpaired t-test, FDR corrected across the shown duration, p< .05). The differences are observed at important points in the phrase structure. Just before the within-phrase transition, Group PS's response latency is shifted relative to Group C (left side of the thin dashed lines). Afterwards, Group PS's response amplitude is lower (right side of the thin dashed lines).

Table 1.

Structure of miniature languages. The two languages had the same vocabulary and basic sentence structure (ABCDEF). Complex sentences were created by applying transformations to pairs of words. In the phrase-structure language (PS: top row), transformations were applied to consistent pairs of words, creating a distributional cue to phrase structure: high transition probabilities within phrases and low transition probabilities across phrase boundaries. In the control language (C: bottom row), transformations were applied to any adjacent pair of words or to AF (to ensure that the statistical distributions of A and F were similar to those of the other adjacent form classes). This created a pattern of relatively flat transition probabilities and no cues to a phrase structure. All sentences in both languages were 4, 6, or 8 words long.

Word pairs for transformations	Example complex sentences		Transition probabilities for basic sentence structure
AB, CD, EF	ABEF	mer$_A$ lev$_B$ ker$_E$ nav$_F$	
	EFCDCD	rud$_E$ sib$_F$ tid$_C$ lum$_D$ rel$_C$ zor$_D$	
	CDEFABCD	tid$_C$ zor$_D$ ker$_E$ nav$_F$ daz$_A$ neb$_B$ sot$_C$ lum$_D$	
AB, BC, CD,	CFDE	jes$_C$ sib$_F$ zor$_D$ taf$_E$	
DE, EF, AF	AFDECD	mer$_A$ sib$_F$ lum$_D$ rud$_E$ rel$_C$ zor$_D$	
	BCADEFAB	lev$_B$ tid$_C$ kof$_A$ zor$_D$ ker$_E$ sib$_F$ mer$_A$ neb$_B$	