NeuroImage 251 (2022) 118979

Contents lists available at ScienceDirect

NeuroImage

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

Delta-band neural activity primarily tracks sentences instead of semantic properties of words

Yuhan Lu[a,b], Peiqing Jin[a], Xunyi Pan[c], Nai Ding[a,b,*]

[a] Department of Neurology, First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310000, China
[b] Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou 310027, China
[c] School of International Studies, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Keywords:
Speech
Grouping
Chunk
Semantics
EEG
Language

ABSTRACT

Human language is generally combinatorial: Words are combined into sentences to flexibly convey meaning. How the brain represents sentences, however, remains debated. Recently, it has been shown that delta-band cortical activity correlates with the sentential structure of speech. It remains debated, however, whether delta-band cortical tracking of sentences truly reflects mental representations of sentences or is caused by neural encoding of semantic properties of individual words. The current study investigates whether delta-band neural tracking of speech can be explained by semantic properties of individual words. Cortical activity is recorded using electroencephalography (EEG) when participants listen to sentences repeating at 1 Hz and word lists. The semantic properties of individual words, simulated using a word2vec model, predict a stronger 1 Hz response to word lists than to sentences. When listeners perform a word-monitoring task that does not require sentential processing, the 1 Hz response to word lists, however, is much weaker than the 1 Hz response to sentences, contradicting the prediction of the lexical semantics model. When listeners are explicitly asked to parse word lists into multi-word chunks, however, cortical activity can reliably track the multi-word chunks. Taken together, these results suggest that delta-band neural responses to speech cannot be fully explained by the semantic properties of single words and are potentially related to the neural representation of multi-word chunks.

1. Introduction

It remains debated whether the brain can group individual items in a sequence into chunks and builds a unified neural representation for each chunk in processing sequences. In the domain of speech processing, it is elusive how the brain integrates information across words to understand sentences and passages (Goucha et al., 2017; Hagoort and Indefrey, 2014; Pylkkänen, 2019). At one end of the spectrum, it has been hypothesized that the brain applies a set of syntactic rules to recursively combine words into larger chunks, forming a hierarchically organized syntactic structure, and then derives meaning of the sentence based on its syntactic structure (Chomsky, 1957; Frazier and Fodor, 1978; Friederici, 2002). At the other end of the spectrum, it has been hypothesized that the brain does not construct multi-word chunks at all, but directly integrates information across words by statistical and semantic analysis instead (Elman, 1990; Frank et al., 2012).

Neuroscientific studies have also been used to test whether the brain groups words into chunks, i.e., phrases, during on-line speech comprehension, and the results are mixed. Some studies are in favor of the chunking hypothesis. For example, an EEG closure positive shift (CPS) response is observed at phrasal boundaries (Li and Yang, 2009; Steinhauer et al., 1999), suggesting that the brain is sensitive to phrasal boundaries. More recent studies have demonstrated low-frequency EEG and magnetoencephalography (MEG) activity can track the time course of phrases and sentences (Ding et al., 2016, 2018; Keitel et al., 2018; Kulasingham et al., 2021), which has been viewed as evidence that the brain applies syntactic rules to group events into superordinate chunks (Ding et al., 2017a; Martin and Doumas, 2017; Meyer and Gumbert, 2018). Furthermore, computational models that incorporate phrasal structures can predict the neural responses to language (Artoni et al., 2020; Brennan and Hale, 2019; Fedorenko et al., 2016; Hale et al., 2018; Nelson et al., 2017). Another line of research, however, argues that models that lack phrasal structure can better predict the EEG response to language (Frank et al., 2015) and it has been proposed that neural tracking of phrases and sentences may potentially be explained by the semantic properties of individual words alone (Frank and Yang, 2018). For example, since the sentences used in Ding et al. (2016) have a regular syntactic structure, i.e., adjec-

---

[*] Corresponding author at: Department of Neurology, First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310000, China.
E-mail address: ding_nai@zju.edu.cn (N. Ding).

https://doi.org/10.1016/j.neuroimage.2022.118979.
Received 26 September 2021; Received in revised form 29 January 2022; Accepted 6 February 2022

tive + noun + verb + noun, the presentation rate of adjectives and verbs was identical to the presentation rate of sentences. Consequently, it is possible that the neural activity appearing to track sentences tracks lexical semantical properties that differentiate adjectives and verbs as a matter of fact.

To tease apart the role of syntactic structure and lexical semantical properties, a recent study presents sequences of words that do not construct sentences and explicitly asks participants to apply an artificial rule to chunk the sequences (Jin et al., 2020). It is found that cortical activity tracks multi-word chunks defined by the explicit chunking rule, instead of lexical semantical properties. The study demonstrates that neural activity can track multi-word chunks during an explicit sequence chunking task. It remains elusive, therefore, to what extent the neural tracking of sentences can be explained by lexical semantical properties. It is also elusive whether neural activity will predominantly track lexical semantical properties when listeners attend to the semantic information of words instead of performing a sequence chunking task.

Here, we investigate how task and sentential structure separately influence neural tracking of word sequences. We use a word-monitoring task to direct listeners' attention to semantic properties of individual words. Under this task, to investigate the influence of sentential structures, we present to the listeners both sentence sequences and word lists that are designed to frequency tag the neural responses tracking lexical semantical properties. We also predict the neural response tracking lexical semantical properties using the word2vec model, a connectionism model that describes semantic relationship between words (Bengio et al., 2000; Mikolov et al., 2013), and test whether the measured neural responses are consistent with the prediction of the model. Furthermore, to analyze to what extent the neural response is modulated by the task, we also compare the neural responses recorded in the word-monitoring task with the responses in an explicit sequence chunking task in which listeners employ a sequence chunking rule to parse word lists into multi-word chunks (Jin et al., 2020).

2. Materials and methods

2.1. Participants

Thirty-two native listeners of Mandarin Chinese (20–30 years old, mean 23 years old, 17 males) participated in the study. All participants were right-handed, with no self-reported hearing loss or neurological disorders. The experimental procedure was approved by the Research Ethics Committee of the College of Medicine, Zhejiang University (2019-047) and was in accordance with the declaration of Helsinki. All participants provided written informed consent before participating in the experiment and received financial compensation in cash after the experiment. The payment was calculated as the sum of base pay (¥60) plus bonus (¥30 × task accuracy).

2.2. Stimuli

The study presented two kinds of word sequences, i.e., word list and sentence sequence (Fig. 1A). Each sequence consisted of 24 disyllabic words in Mandarin Chinese. All disyllabic words were synthesized independently using the iFLYTEK synthesizer (http://peiyin.xunfei.cn/; Mandarin Chinese; female voice, Xiaoying). The intensity of each word was adjusted to be the same and the duration was adjusted to 500 ms, following the procedure in Ding et al. (2016). Within a word, however, no additional control was applied to adjust the intensity and duration of individual syllables, so that coarticulation may exist between syllables (c.f., Levelt, 1993), i.e., the syllables within a word are not acoustically independent. The disyllabic words synthesized as a whole sounded more natural, compared with speech materials in which each syllable was independently synthesized. The synthesized disyllabic words were directly concatenated to construct sequences, and no acoustic gaps were inserted between words. Therefore, each disyllabic word was an acoustically independent unit and was isochronously presented at 2 Hz. The spectrum of stimuli intensity only showed that acoustic fluctuations at 2 Hz (P = 0.0002 for both word lists and sentence sequences; paired two-sided bootstrap, FDR corrected; Fig. A.1A).

In the sentence sequence condition, 80 4-syllabe sentences were constructed, in which the first two syllables formed a noun (or a common noun phrase) and the last two syllables formed a verb (or a common verb phrase) (Fig. 1A). The sentences were all common sentences in Chinese (Table A.1) and on average each sentence appeared 5.7 times in the experiment. Unlike sentence sequences, word lists presented no syntactic structure but were constructed with disyllabic nouns (N = 240), either living (l) or nonliving (n) nouns. Living nouns included two subcategories, i.e., animals (N = 60; e.g., monkey, panda) and plants (N = 60; e.g., tulip, strawberry). In each word list, each living noun was randomly selected from one subcategory. Likewise, nonliving nouns included two subcategories, i.e., small manipulatable objects (N = 60; e.g., teacup, toothbrush) and large non-manipulatable objects (N = 60; e.g., playground, hotel). Each nonliving noun in a word list was randomly selected from one subcategory. In a word list, the semantic categories of words changed cyclically with a period of two words (Fig. 1A), and each word was selected independent of its neighboring words.

Since living and nonliving nouns formed two distinct categories, the lexical semantical properties alternate periodically with periods of two words. Accordingly, neural activity tracking the lexical semantical properties of words was expected to show periodicity at 1 Hz, which was further illustrated using word2vec-based model simulation (see *Results*). Additionally, to build outlier trials, a pool of 42 disyllabic abstract nouns (e.g., honor, spirit) was used and each abstract noun only appeared once in the experiment.

2.3. Procedures and tasks

At the beginning of the experiment, participants were required to familiarize themselves with the synthesized words. In this session, all words used in this study were presented to participants in a random order. When participants heard a word, they pressed a key to see the word on computer screen. Then, they could choose to listen to the word again or proceed to the next word by pressing different keys.

In the experiment, neural responses were recorded using EEG. The experiment consisted of three conditions that were presented in separate blocks. Two conditions presented word lists and sentence sequences while the participants performed a word-monitoring task, and another condition presented word lists while the participants performed a sequence chunking task. Behavioral performance was reported in Table A.2. In each block, 30 normal sequences and 7 outlier sequences were mixed and presented in a random order. Participants had a rest between blocks.

The first and the second conditions presented word lists and sentence sequences, respectively, and participants were asked to perform the word-monitoring task. In this task, participants were instructed to detect occasionally presented abstract nouns (Fig. 1B). An outlier trial had the same design as a normal trial did, except that two words in non-adjacent sites were replaced with two abstract nouns in sequences.

Note that before tasks of the first two conditions began, participants were told that normal word lists and sentence sequences only contained concrete words, but were not informed of the variations in sequence structure, that is, the cyclic alternation of nouns of different semantic categories in word lists and “noun + verb” combination in sentence sequences. Moreover, the first condition was further separated into two blocks: one block consisting of sequences starting with a living noun, and the other starting with a nonliving noun. In data analysis, the two blocks were separately analyzed, and the power spectra were averaged across the two blocks.

The third condition presented word lists, and participants performed a sequence chunking task with rules explicitly explained to them. Par-

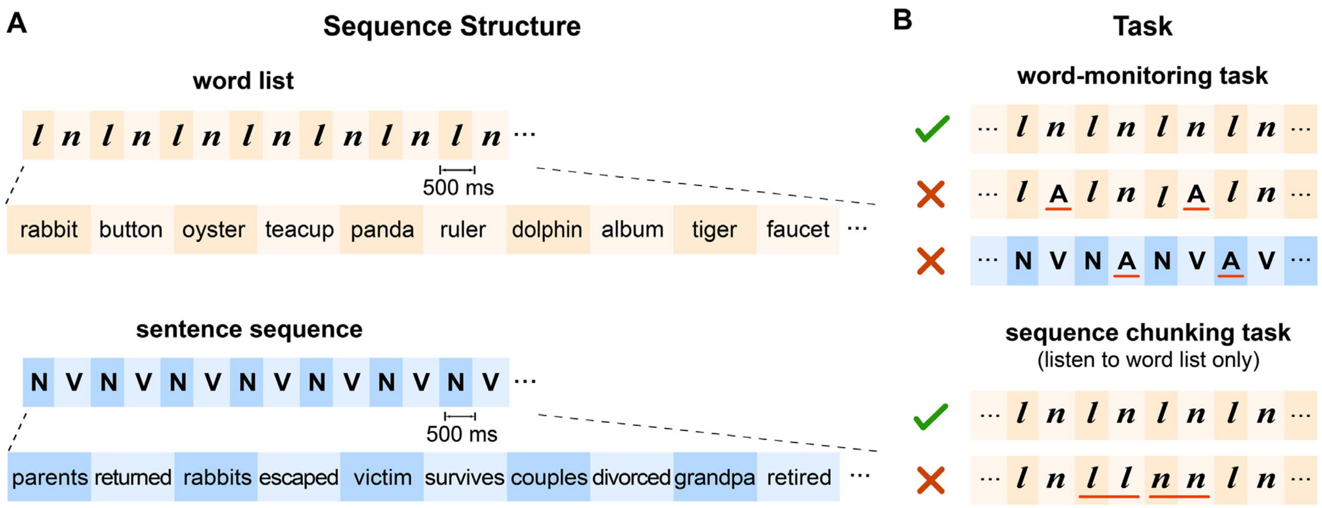

Fig. 1. Stimuli and tasks. ***A,*** A word list consists of isochronously-presented disyllabic nouns from two semantic categories, i.e., living (*l*) and nonliving (*n*) nouns. In the word list, the semantic categories of words form cyclic patterns repeating every 2 words. In a sentence sequence, each sentence is constructed with a disyllabic noun (N) followed by a disyllabic verb (V). ***B,*** During the word-monitoring task, outlier trials are constructed by replacing two words in non-adjacent sites with two abstract nouns (A), and participants are instructed to detect these abstract nouns while listening to word lists and sentence sequences. During the sequence chunking task, outlier trials are constructed by switching the position of a living noun with its neighboring nonliving noun in word lists, thus two words in a chunk fall into the same semantic category forming an invalid chunk. Participants are instructed to detect invalid chunks while listening to word lists.

ticipants were told that two neighboring words in the word lists could construct a chunk on the condition and the two words fell into different semantic categories, i.e., living nouns and nonliving nouns. During the task, they had to detect invalid chunks that contained words of the same category (Fig. 1B). To build such invalid chunks in an outlier trial, a living noun was switched with its neighboring nonliving noun, and therefore two nouns in a chunk fell into the same semantic category, violating the chunk construction rule. The sequences in this condition started with either a living noun or a nonliving noun, with equal probability.

In the experiment, participants were not explicitly informed of the sequence structure in the first and second condition. The order of the first and the second conditions were counterbalanced, while the third condition was always administered at the end. After listening to a sequence in a trial, participants pressed different keys to indicate whether it was a normal or an outlier sequence. Then the next sequence was presented after a silent interval randomized between 1 and 2 s (uniform distribution). Only trials with normal sequences were used in EEG analysis.

Before EEG recording, participants were familiarized with the word-monitoring task by listening to sample word lists and sentence sequences. Before the sequence chunking task, the EEG session was given a pause when participants were instructed about the sequence structure of word lists. After receiving the instruction, the participants were given two normal sequences and two outlier sequences for familiarization. When listening to the outlier sequences, they were asked to verbally report the invalid chunks as soon as they heard them. The sequences could be replayed upon request. The participants then went through a practice session; it ended after the participants made 8 correct responses in 10 consecutive sequences. The EEG experiment resumed after the practice session.

2.4. Data acquisition

In the experiment, EEG responses were recorded from 5 channels, i.e., Cz, Fz, FCz, FC3, and FC4, using a 64-channel Biosemi ActiveTwo system at the Zhejiang University. The 5 channels were chosen since previous studies showed that they were sensitive to neural tracking of spoken sentences (Ding et al., 2018; Jin et al., 2018). Additionally, four electrodes were used to record horizontal and vertical EOGs. Two reference electrodes were placed at the left and right mastoids and their average was the reference for EEG.

2.5. Data processing

Only the neural responses to normal sequences were analyzed. To remove ocular artifacts in EEG, the horizontal and vertical EOG were regressed out using the least-squares method (Ding et al., 2017b). To show broadband responses between 0 and 45 Hz (Fig. A.1B), the EEG signals were first down-sampled at 512 Hz. Since the study focused on responses at 1 and 2 Hz, the EEG signals were further bandpass filtered between 0.3 and 2.7 Hz using a linear-phase finite impulse response (FIR) filter (-6 dB attenuation at the cut-off frequencies, 10-s Hamming window), and down-sampled at 20 Hz. The narrowband response to each sequence was extracted and was referred to as a trial.

2.6. Frequency-domain analysis

In frequency-domain analysis, to avoid the influence of onset response, the response during the first two seconds of each trial were removed. Consequently, the neural response was 10 s in duration for each trial. The average of all trials was transformed into the frequency domain using the Discrete Fourier Transform (DFT) without any additional smoothing window. The frequency resolution of the DFT analysis was 1/10 Hz. If the complex-valued DFT coefficient at frequency f was denoted as $X(f)$, the response power was $|X(f)|^2$, The DFT was separately applied to each EEG channel. In EEG power analyses, neural response power was averaged across channels.

2.7. Model simulations

2.7.1. Single-Word semantic model

The study was designed to investigate whether the lexical semantics model proposed by Frank and Yang (2018) can fully explain the low-frequency neural response to sentences. The model by Frank and Yang assumes that neural activity only encodes the distributive semantic features of individual words, and it is implemented using the following procedure.

Pulse Sequence: We constructed the lexical semantics model to simulate how the semantical properties of individual words contributed to the neural response to word lists and sentences. In the model, the smallest unit being considered was word. The semantical properties of individual words were first simulated using a pulse sequence, in which a pulse was placed at the onset of each word. The pulse amplitude of words was determined by the word2vec representation (Bengio et al., 2000; Li et al., 2018). Each word was represented by 300 feature dimensions, which were learned based on a large corpora (the 'combination' corpora in Li et al., 2018). Each of the 300-feature dimension was coded by a real number and modulated the amplitude of the pulse. The neural response to each word was simulated by 300 pulse sequences.

Neural Response Waveform Simulation: Neural response waveforms were further simulated by convolving the pulse sequences with a 500-ms duration Gaussian window. For the lexical semantics model, the neural response to each one of the 300 features was independently simulated and transformed into frequency domain. In fact, the experiment only recorded from 5 electrodes and therefore at most could resolve 5 feature dimensions. However, to assume the highest power of the lexical semantics model, we used all the 300 feature dimensions in the simulation. The power spectrum was averaged over feature dimensions (Frank and Yang, 2018).

2.7.2. *Other models*

Although the purpose of the study was to test the Frank and Yang model, we also simulated the properties of neural responses tracking other features of speech (Fig. A.2). The first model considering semantic correlation between consecutive words in a word sequence, i.e., semantic relatedness model. Different from the lexical semantics hypothesis, it was certainly possible that the neural encoding of semantic relatedness (i.e., correlation) between individual words in a sentence could explain the mental representation of the sentence without inferring sentential processing. The semantic relatedness model was built on the lexical semantics model (Broderick et al., 2018) and was characterized by the correlation coefficient between the current word and the adjacent previous word. The correlation coefficient was always a scalar. Since the neural response to a stimulus was usually weaker instead of stronger if the stimulus was preceded by a similar stimulus, we used 1 minus the correlation coefficient to modulate the amplitude of each pulse.

The second and the third model separately considered the occurrence frequency of individual words and the bigram probability. We retrieved the occurrence frequency of each word and each word pair based on a large Chinese corpus (Zhan et al., 2019). For occurrence frequency model and bigram probability model, the amplitude of the pulse to each word was set to its value of occurrence frequency and bigram probability, respectively.

2.8. *Statistical tests*

All tests were carried out using bias-corrected and accelerated bootstrap (Efron and Tibshirani, 1994). In the bootstrap procedure, data of all participants was resampled 10,000 times with replacement. All comparisons in this study were paired comparisons. For one-sided comparison, if the data in one condition was greater than that in the other condition in A% of the resampled data, the significance level was (100A + 1)/10,001. For two-sided comparisons, if the data was greater in one condition for A% of the resampled data, the significance level was (200A + 1)/10,001. A false discovery rate (FDR) correction was applied to adjust the *P*-value during multiple comparisons.

2.8.1. *Spectral peak*

The statistical significance of a spectral peak at frequency *f* was tested by comparing the response power at *f* with the response power in the frequency bin just below *f* (Jin et al., 2018). The comparison was one-sided. The comparison only considered the frequency bin below *f* since spontaneous neural activity had a 1/f spectrum and the response power in the frequency bin above *f* tended to be weaker than the power at *f* even without stimulus-evoked activity. This significance test was only applied to the response power at 1 and 2 Hz.

2.8.2. *Response power*

The comparisons in normalized power between conditions were two-sided. The normalized response power was defined as the difference between the power at *f* and the power in the frequency bin just below *f*. This significance test was only applied to the response power at 1 and 2 Hz.

2.8.3. *Post-hoc effect size calculation*

To validate the sample size in the study was appropriate, we calculated effect size and power to the 1 Hz spectral peak using G*Power Version 3.1 (Faul et al., 2007). We applied a paired *t*-test to compare the power at 1 Hz and its neighboring frequency bin just below 1 Hz. We calculated effect size *d* and power based on their comparison. For the effect size observed in the data set, the study was powerful with the described sample population and the α level of 0.05 (Table A.3).

3. Results

3.1. *Model predictions*

We simulated the neural responses to word lists and sentence sequences based on the word2vec-based lexical semantics model. The lexical semantics model assumed that some components of the EEG responses were tuned to semantic features which were captured using the word2vec model. Through statistical analysis of large corpus, the word2vec algorithm constructed a 300-dimensional vectorial representation for each word that could characterize its lexical semantic properties. In general, words with similar meanings had similar word2vec representations. For the words used in the current study, the correlation between word2vec representations was higher for words from the same subcategory, e.g., animals and plants (Fig. 2A). In each word list, living nouns were selected from a fixed subcategory; nonliving nouns were selected from a fixed subcategory. A distinct semantic contrast between living and nonliving nouns was formed, and therefore the word2vec representation could capture such a contrast between the nouns in a word list. In addition, in each sentence sequence, a noun and its following verb established a predication structure indicating obvious syntactic relations between the two words; it could be predicted that the word2vec representation could also capture semantic coherence of words in a sentence sequence. We quantified the semantic correlation between words in both word lists and sentence sequences, and found that the correlation between nouns and verbs in sentence sequences was stronger than that between living and nonliving nouns in word lists (Fig. A.2A; $P = 1 \times 10^{-4}$; paired two-sided bootstrap, FDR corrected).

The lexical semantics model predicted significant 1- and 2 Hz responses to both word lists and the sentence sequences (Fig. 2B; $P = 10^{-4}$ for both responses and frequencies; paired one-sided bootstrap, FDR corrected). Importantly, the lexical semantics model predicted that the 1- and 2 Hz responses to word lists were stronger than those to sentence sequences (Fig. 2C; $P = 0.0003$ for both frequencies; paired two-sided bootstrap, FDR corrected).

Although the purpose of the study was to test the lexical semantics model, to fully characterize the properties of the stimulus sequences, we also analyzed how the occurrence frequency of individual words, bigram probability, and semantic relatedness between consecutive words might differ between the word lists and sentences. The occurrence frequency of words predicted a 1 Hz response to both word lists and sentences (Fig. A.2B; $P = 1 \times 10^{-5}$ for both sequences; paired two-sided bootstrap, FDR corrected), but the power of the 1 Hz response was not significantly different between conditions ($P = 0.274$; paired two-sided bootstrap, FDR corrected). The bigram probability model and semantic relatedness model predicted a significant 1 Hz response to sentences

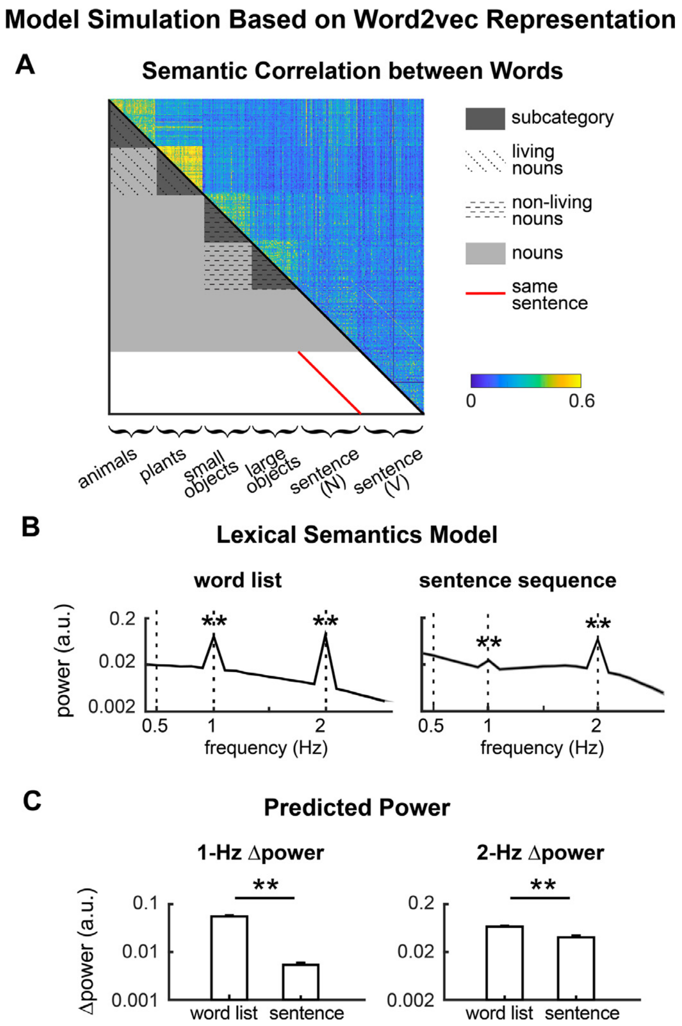

Fig. 2. Model simulations. ***A***, Semantic correlation between the word2vec representations of words. The plot above the diagonal shows the actual correlation and the plot below the diagonal illustrates regions expected to show higher correlation by different word relations. Words in each subcategory, e.g., animals and plants, are correlated, and words in the same sentence also showed correlation. ***B***, Simulated response spectrum. The lexical semantics model predicts 1- and 2 Hz responses to word lists and sentence sequences. ***C***, Simulated response power. The lexical semantics model predicts stronger 1- and 2 Hz responses to word lists than to sentence sequences. ** $P < 0.005$.

only ($P = 1 \times 10^{-5}$ and 1×10^{-4}, respectively; paired two-sided bootstrap, FDR corrected; Fig. A.2C and A.2D), but not to word lists.

Next, we tested whether the actual neural responses were consistent with the predictions of the lexical semantics model using an EEG experiment. The experiment tested whether the actual neural response to word lists had a significant 1 Hz component, which was predicted by both the lexical semantics model and the word occurrence frequency model, and whether the 1 Hz response to word lists was stronger than that to sentences, which was predicted by the lexical semantics model alone.

3.2. EEG responses

The experiment presented word lists and sentence sequences to participants, and recorded their neural responses using EEG. Participants were asked to perform a word-monitoring task to detect occasionally-presented abstract words in a sequence. During this task, the sequence structure of word lists and sentence sequences were not disclosed to the participants. In the EEG response, statistically significant 1 Hz responses to word lists and to sentence sequences were observed during the word-monitoring task (Fig. 3A; $P = 0.002$ and $P = 1 \times 10^{-4}$, respectively; paired one-sided bootstrap, FDR corrected), consistent with the prediction of the lexical semantics model and the word occurrence frequency model. Importantly, we found that the 1 Hz response to sentence sequences was significantly stronger than that to word lists (Fig. 3B; $P = 0.001$; paired two-sided bootstrap, FDR corrected), suggesting that the 1 Hz response to sentence sequences cannot be sufficiently explained by the lexical semantics model, which predicted a stronger 1 Hz response to the word lists than to sentence sequences.

We further tested whether the 1 Hz response to word lists could be enhanced by the sequence chunking task in which participants received explicit explanation of the sequence structure of word lists and were required to detect invalid chunks in a sequence. During this task, the 1 Hz response to word lists was significant (Fig. 3A; $P = 10^{-4}$; paired one-sided bootstrap, FDR corrected), and was stronger than that during the word-monitoring task (Fig. 3B; $P = 0.0006$; paired two-sided bootstrap, FDR corrected). This result suggested that implicit syntactic processing and explicit sequence chunking can both drive cortical responses more effectively than lexical semantical properties of single words. Additionally, to evaluate whether the neural response significantly varied over the course of experiment, we separately showed the EEG spectrum for three 10-trials blocks in each condition (Fig. A.3A). The results showed that only the 1 Hz response to sentences decreased over trials: The response in the last 10 trials was significantly weaker than the first 10 trials ($P = 0.015$; paired two-sided bootstrap, FDR corrected; Fig. A.3B).

Significant 2 Hz responses in all conditions were observed (Fig. 3A; $P = 0.0003$ and $P = 10^{-4}$ for word lists and sentence sequences, respectively, during the word-monitoring task, and $P = 10^{-4}$ for word lists during the sequence chunking task; paired one-sided bootstrap, FDR corrected). The 2 Hz response to sentence sequences was weaker than that to word lists during both the word-monitoring task and sequence chunking task (Fig. 3B; $P = 0.011$ for both comparisons; paired two-sided bootstrap, FDR corrected).

4. Discussion

It has been debated for a long-time whether the brain constructs phrasal and sentential representations during speech comprehension. Recently, it has been argued that cortical activity tracks sentential and phrasal structures in speech, demonstrating sentential and phrasal representations in the brain (Ding et al., 2016). Nevertheless, it has been argued that apparent sentential/phrasal tracking is caused by neural encoding of the lexical properties of individual words (Frank and Yang, 2018). The current study shows that during a word-monitoring task, the EEG response to word lists shows a 1 Hz response (Fig. 3A), which is predicted by both a lexical semantics model and a word occurrence frequency model (Figs. 2B and A.2B). The lexical semantics model further predicts a stronger 1 Hz response to word lists than sentences (Fig. 2C), but the 1 Hz neural response to word lists detected in reality is significantly weaker than that to sentences (Fig. 3B). This suggests that the semantic properties of individual words alone cannot fully explain neural tracking of sentences. The cortical activity, however, can strongly track multi-word chunks defined by an artificial chunking rule during an explicit sequence chunking task, and the magnitude of the response to the multi-word chunks is similar to that to sentence sequences in the word-monitoring task.

4.1. Neural encoding of lexical semantic properties

The lexical semantics model is constructed on the basis of the findings that words from different grammatical categories, e.g., verbs and nouns, are separately represented in the brain (Vigliocco et al., 2011; Yang et al., 2017) and can be selectively impaired (Caramazza and

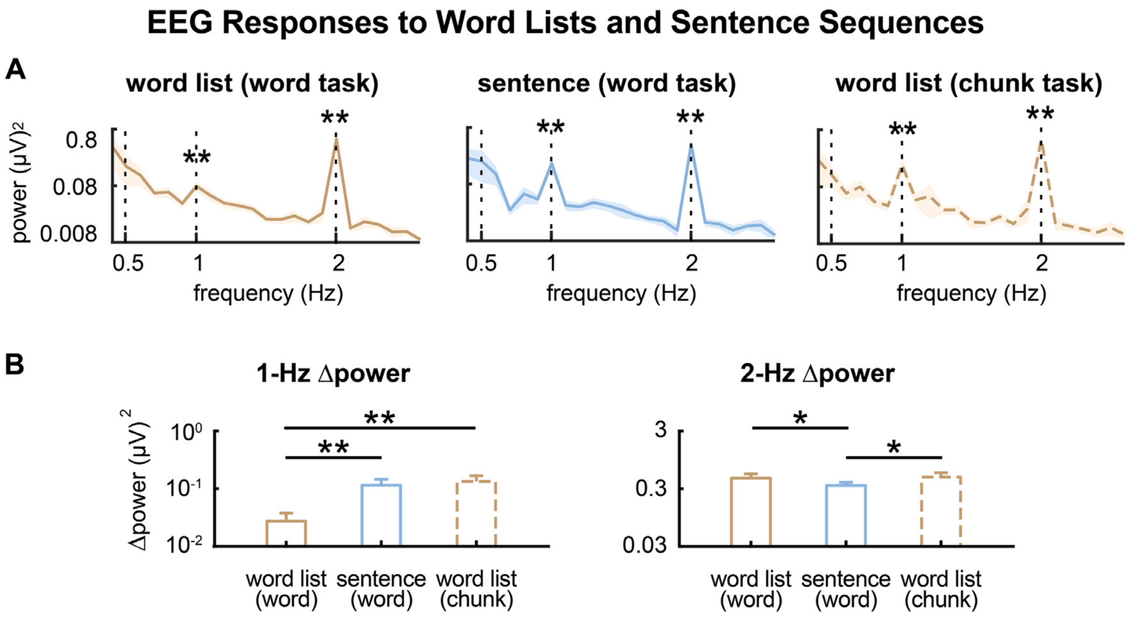

Fig. 3. EEG responses to word lists and sentence sequences. ***A***, The EEG response spectrum averaged over participants and channels show 1 Hz and 2 Hz response peaks. The shaded area covers 1 SEM over participants on each side. ***B***, Response power at 1 and 2 Hz. For word lists, the 1 Hz response is stronger during the explicit chunking task than that during the word-monitoring task. During the same word-monitoring task, the 1 Hz response to sentence sequences is stronger than that to word lists. The 2 Hz response to sentences was weaker than that to word lists in both the word-monitoring task and the sequence chunking task. * $P < 0.05$, ** $P < 0.005$.

Hillis, 1991; Daniele et al., 1994). Similarly, words from distinct categories, e.g., living and nonliving nouns, are separately represented in the brain and can be selectively impaired (Bi et al., 2016; Warrington and Shallice, 1984). A number of studies have suggested that the semantic categories of individual words can be detected with the spatial resolution of MEG recoding. For example, using a multivariate analysis of multi-channel MEG decoding, it has been shown that the neural responses to living and nonliving nouns can be distinguished (Chan et al., 2011). Using similar decoding approaches, the semantic categories of visually presented objects can be successfully decoded (Carlson et al., 2011; Sudre et al., 2012).

Although these previous studies have shown that the spatial pattern of cortical activity carries semantic information using multivariate neural decoding approaches, to our knowledge no study has shown that univariate MEG/EEG responses (e.g., single-channel responses or global field power) could clearly distinguish the semantic properties of individual words. The delta-band neural response tracking sentences, however, can be observed in single MEG/EEG sensors and in the global field power (Ding et al., 2016; Jin et al., 2018), and the sentence-rate response is typically as strong as the word-rate response. More importantly, the lexical semantics model fails to predict a stronger 1 Hz response to sentence sequences than to word lists.

Although the word semantics model predicts equal 1 Hz responses to sentences and word lists, only sentences contain a regular 1 Hz change in part-of-speech information, which can also potentially drive a 1 Hz neural response. Two recent studies, however, provide negative evidence for neural tracking of part-of-speech information (Burroughs et al., 2021; Lo, 2021). For example, a study by Lo (2021) reverses the order of every pair of words, turning normal four-syllable sentences (e.g., cotton sheep eat grass) into word lists (e.g., sheep cotton grass eat). For the word lists constructed this way, part-of-speech information keeps varying periodically but no neural response is observed at the frequency of the part-of-speech information alternation. The study by Burroughs et al. (2021) compares the responses to sequences of two-word adjective + noun phrases (e.g., cold food) and sequences of adjective + verb word pairs (e.g., rough give). The part-of-speech information alternates in both kinds of sequences but a neural response tracking the bi-word rhythm is only observed in the phrase condition, suggesting that part-of-speech information alone cannot drive neural tracking of word pairs.

4.2. Neural encoding of context-dependent properties of words

The neural response to a word, e.g., the N400 component, is sensitive to the context, such as the semantic relation to previous words (Kutas and Federmeier, 2011; Kutas and Hillyard, 1980, 1984). The N400 response reduces throughout the time course of a normal sentence (Halgren et al., 2002; Kutas and Federmeier, 2011; Lau et al., 2008; Roehm et al., 2004), which can result in a sentence-tracking response. A factor similar to semantic relatedness is lexical surprisal/predictability. Recent studies show that neural activity can track the lexical surprisal of words (Gillis et al., 2021) independent of their semantic relatedness (Frank et al., 2015; Willems et al., 2015). In the current study, the bi-gram probability calculated from large corpora shows a stronger 1 Hz change for sentences than word lists. Furthermore, each sentence was presented 5, 6 times during the experiment. In contrast, the living and nonliving words in word lists were randomly paired and word pairs were seldom repeated. Therefore, participants could also learn online that later words in a sentence are more predictable than the words in word lists.

Therefore, semantic relatedness between words and word surprisal/predictability can both predict a stronger 1 Hz response to sentences than to word lists. However, these measures fail to predict the 1 Hz response to word lists, and cannot explain task modulation of the 1 Hz response. The purpose of the current study is to investigate whether the distributive semantic properties of individual words could fully explain delta-band EEG responses, as is proposed by Frank and Yang (2018). Therefore, the stimuli are not optimized to isolate the effects of, e.g., the semantic relatedness between words and lexical surprisal/predictability. Future studies are needed to analyze how these factors contribute to neural tracking of speech.

4.3. Neural encoding of multi-word chunks

It has been proposed that the brain can group multiple words into a chunk and builds a neural representation for the chunk as a whole. On the one hand, during normal language processing, words may be implicitly integrated into chunks based on syntactic rules. On the other hand, here we employ a task to ask participants to group words into chunks by applying an explicit sequence chunking rule. It should be noted that the explicit chunking task employed here could in principle be completed using the same strategy in the word monitoring task, i.e., monitoring the appearance of living or nonliving words. If participants used the same strategy in two tasks, however, their neural responses would not have differed between the word-monitoring and sequence chunking conditions. Furthermore, there is also evidence suggesting that chunking can occur spontaneously (Wymbs et al., 2012). Humans prefer to group items into two- to four-items chunks (Lerdahl and Jackendoff, 1981; Miller, 1956). However, spontaneous grouping cannot easily explain why sentences and word lists elicit different responses. Last, it can be argued that even during the word monitoring task the participants implicitly learn the sequence structure. However, even if the participants learn that the sequence alternates between two classes of words, it is unclear whether they parse the sequence into units of *nl, ln, nlnl, lnln*, or other chunks. Furthermore, the 1 Hz response to word lists during the word monitoring task does not increase over time (Fig. A.3B) and therefore is not likely to be the consequence of learning.

Chunking can be implemented through multiple mechanisms. For example, it has been proposed that chunking can be implemented using either automatic processes or more controlled processes, which engage different neural circuits (Jeon and Friederici, 2015). However, it is also possible that there is a common core cortical area for chunking in general. For example, functional MRI studies show that ventrolateral prefrontal cortex, including the Broca's area, not only is a core area for language processing, but also can be activated by rule-based nonlinguistic sequential processing tasks (Koechlin and Jubault, 2006; Thompson-Schill et al., 2005). Furthermore, chunking can also be driven by different cues and occur on multiple dimensions. For example, in natural speech, prosody is a critical factor to drive parsing and an EEG CPS response is observed at prosodic phrasal boundaries (Li and Yang, 2009; Steinhauer et al., 1999). The current speech materials, however, are deprived of chunk-related prosodic cues. Furthermore, chunking during language comprehension can also occur in semantic and syntactic dimensions (Artoni et al., 2020; Brennan et al., 2012; Nelson et al., 2017; Zhang and Pylkkanen, 2015, 2018), and future studies are needed to investigate which dimension is more strongly reflected in the speech-tracking neural responses.

5. Conclusions

In sum, the current study and previous study (Jin et al., 2020) demonstrate that the semantic properties of words are insufficient to explain delta-band cortical response to sentences. In contrast, implicit syntactic processing and explicit sequence chunking can more effectively drive delta-band cortical responses. These results suggest that delta-band neural activity is better explained by a chunk-level neural representation, instead of lexical-level neural representation.

Declaration of Competing Interest

None.

Credit authorship contribution statement

Yuhan Lu: Data curation, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Peiqing Jin:** Software, Validation, Formal analysis, Resources, Writing – original draft. **Xunyi Pan:** Writing – review & editing, Supervision, Funding acquisition. **Nai Ding:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We thank Xing Tian, Stefan L. Frank, Lucia Melloni, Lang Qin, and David Poeppel for thoughtful comments on previous versions of the manuscript, Jiajie Zou for discussion and generating the speech materials, and Jiaxin Gao for editing the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China [grant number: 2021ZD0204105], the National Natural Science Foundation of China [grant number: 31771248, ND] and the Zhejiang Provincial Natural Science Foundation of China [grant number: LY20C090008, XP].

Data and code availability

All data and code that support the findings of this study are available at: Lu, Yuhan; Ding, Nai (2021), "Delta-band neural activity primarily tracks sentences instead of semantic properties of words", Mendeley Data, v1 http://dx.doi.org/10.17632/dhn66g7fgw.1

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.118979.

References

Artoni, F., d'Orio, P., Catricala, E., Conca, F., Bottoni, F., Pelliccia, V., Sartori, I., Russo, G.L., Cappa, S.F., Micera, S., Moro, A., 2020. High gamma response tracks different syntactic structures in homophonous phrases. Sci. Rep. 10, 7537. doi:10.1038/s41598-020-64375-9.

Bengio, Y., Ducharme, R., Vincent, P., 2000. A neural probabilistic language model. J. Mach. Learn. Res. 3, 932–938. doi:10.1162/153244303322533223.

Bi, Y., Wang, X., Caramazza, A., 2016. Object domain and modality in the ventral visual pathway. Trends Cogn. Sci. 20, 282–290. doi:10.1016/j.tics.2016.02.002, (Regul. Ed.).

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D.J., Pylkkanen, L., 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. Brain Lang. 120, 163–173. doi:10.1016/j.bandl.2010.04.002.

Brennan, J.R., Hale, J.T., 2019. Hierarchical structure guides rapid linguistic predictions during naturalistic listening. PLoS One 14, e0207741. doi:10.1371/journal.pone.0207741.

Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. 28. doi:10.1016/j.cub.2018.01.080, 803-809 e803.

Burroughs, A., Kazanina, N., Houghton, C., 2021. Grammatical category and the neural processing of phrases. Sci. Rep. 11, 2446. doi:10.1038/s41598-021-81901-5.

Caramazza, A., Hillis, A.E., 1991. Lexical organization of nouns and verbs in the brain. Nature 349, 788. doi:10.1038/349788a0, doi.org/.

Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. J. Vis. 11, 9. doi:10.1167/11.10.9.

Chan, A.M., Halgren, E., Marinkovic, K., Cash, S.S., 2011. Decoding word and category-specific spatiotemporal representations from MEG and EEG. NeuroImage 54, 3028–3039. doi:10.1016/j.neuroimage.2010.10.073.

Chomsky, N., 1957. Syntactic structures. Mouton de Gruyter. 10.2307/2269813

Daniele, A., Giustolisi, L., Silveri, M.C., Colosimo, C., Gainotti, G., 1994. Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. Neuropsychologia 32, 1325–1341. doi:10.1016/0028-3932(94)00066-2.

Ding, N., Melloni, L., Tian, X., Poeppel, D., 2017a. Rule-based and word-level statistics-based processing of language: insights from neuroscience. Lang. Cogn. Neurosci. 32, 570–575. doi:10.1080/23273798.2016.1215477.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017b. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). Front. Hum. Neurosci. 11, 481. doi:10.3389/fnhum.2017.00481.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164. doi:10.1038/nn.4186.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., Zhang, J., 2018. Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. J. Neurosci. 38, 1178–1188. doi:10.1523/JNEUROSCI.2606-17.2017.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press doi:10.1201/9780429246593.

Elman, J., 1990. Finding structure in time. Cogn. Sci. 14, 179–211. doi:10.1207/s15516709cog1402_1.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191. doi:10.3758/BF03193146.

Fedorenko, E., Scott, T.L., Brunner, P., Coon, W.G., Pritchett, B., Schalk, G., Kanwisher, N., 2016. Neural correlate of the construction of sentence meaning. Proc. Natl. Acad. Sci. U. S. A. 113, E6256–E6262. doi:10.1073/pnas.1612132113.

Frank, S.L., Bod, R., Christiansen, M.H., 2012. How hierarchical is language use? Proc. Biol. Sci. 279, 4522–4531. doi:10.1098/rspb.2012.1741.

Frank, S.L., Otten, L.J., Galli, G., Vigliocco, G., 2015. The ERP response to the amount of information conveyed by words in sentences. Brain Lang. 140, 1–11. doi:10.1016/j.bandl.2014.10.006.

Frank, S.L., Yang, J., 2018. Lexical representation explains cortical entrainment during speech comprehension. PLoS One 13, e0197304. doi:10.1371/journal.pone.0197304.

Frazier, L., Fodor, J., 1978. The sausage machine: a new two-stage parsing model. Cognition 6, 291–325. doi:10.1016/0010-0277(78)90002-1.

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6, 78–84. doi:10.1016/S1364-6613(00)01839-8, (Regul. Ed.).

Gillis, M., Vanthornhout, J., Simon, J.Z., Francart, T., Brodbeck, C., 2021. Neural markers of speech comprehension: measuring EEG tracking of linguistic speech representations, controlling the speech acoustics. bioRxiv. 10.1523/JNEUROSCI.0812-21.2021

Goucha, T., Zaccarella, E., Friederici, A.D., 2017. A revival of Homo loquens as a builder of labeled structures: neurocognitive considerations. Neurosci. Biobehav. Rev. 81, 213–224. doi:10.1016/j.neubiorev.2017.01.036.

Hagoort, P., Indefrey, P., 2014. The neurobiology of language beyond single words. Annu. Rev. Neurosci. 37, 347–362. doi:10.1146/annurev-neuro-071013-013847.

Hale, J., Dyer, C., Kuncoro, A., Brennan, J.R., 2018. Finding syntax in human encephalography with beam search. arXiv preprint arXiv:1806.04127.

Halgren, E., Dhond, R.P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J.D., Dale, A.M., 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. NeuroImage 17, 1101–1116. doi:10.1006/nimg.2002.1268.

Jeon, H.A., Friederici, A.D., 2015. Degree of automaticity and the prefrontal cortex. Trends Cogn. Sci. 19, 244–250. doi:10.1016/j.tics.2015.03.003.

Jin, P., Lu, Y., Ding, N., 2020. Low-frequency neural activity reflects rule-based chunking during speech listening. Elife 9. doi:10.7554/eLife.55613.

Jin, P., Zou, J., Zhou, T., Ding, N., 2018. Eye activity tracks task-relevant structures during speech and auditory sequence perception. Nat. Commun. 9, 5374. doi:10.1038/s41467-018-07773-y.

Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. 16, e2004473. doi:10.1371/journal.pbio.2004473.

Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. Neuron 50, 963–974. doi:10.1016/j.neuron.2006.05.017.

Kulasingham, J.P., Joshi, N.H., Rezaeizadeh, M., Simon, J.Z., 2021. Cortical processing of arithmetic and simple sentences in an auditory attention task. J. Neurosci. 41, 8023–8039. doi:10.1523/jneurosci.0269-21.2021.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. doi:10.1146/annurev.psych.093008.131123.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205. doi:10.1126/science.7350657.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163. doi:10.1038/307161a0.

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9, 920–933. doi:10.1038/nrn2532.

Lerdahl, F., Jackendoff, R., 1981. On the theory of grouping and meter. Music. Q. 67, 479–506.

Levelt, W.J., 1993. Speaking: from Intention to Articulation. MIT Press doi:10.7551/mitpress/6393.001.0001.

Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., Du, X., 2018. Analogical reasoning on chinese morphological and semantic relations. arXiv preprint arXiv:1805.06504.

Li, W.J., Yang, Y., 2009. Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. Neuroscience 158, 1416–1425. doi:10.1016/j.neuroscience.2008.10.065.

Lo, C.W., 2021. Testing low-frequency neural activity in sentence understanding. Linguistics doi:10.7302/2952, University of Michigan..

Martin, A.E., Doumas, L.A., 2017. A mechanism for the cortical computation of hierarchical linguistic structure. PLoS Biol. 15, e2000663. doi:10.1371/journal.pbio.2000663.

Meyer, L., Gumbert, M., 2018. Synchronization of electrophysiological responses with speech benefits syntactic information processing. J. Cogn. Neurosci. 30, 1066–1074. doi:10.1162/jocn_a_01236.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63, 81. doi:10.1037/h0043158.

Nelson, M.J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S.S., Naccache, L., Hale, J.T., Pallier, C., Dehaene, S., 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. Proc. Natl. Acad. Sci. U. S. A. 114, E3669–E3678. doi:10.1073/pnas.1701590114.

Pylkkänen, L., 2019. The neural basis of combinatory syntax and semantics. Science 366, 62–66. doi:10.1126/science.aax0050.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., Haider, H., 2004. Fractionating language comprehension via frequency characteristics of the human EEG. Neuroreport 15, 409–412. doi:10.1097/00001756-200403010-00005.

Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. 2, 191–196. doi:10.1038/5757.

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., Mitchell, T., 2012. Tracking neural coding of perceptual and semantic features of concrete nouns. NeuroImage 62, 451–463. doi:10.1016/j.neuroimage.2012.04.048.

Thompson-Schill, S.L., Bedny, M., Goldberg, R.F., 2005. The frontal lobes and the regulation of mental activity. Curr. Opin. Neurobiol. 15, 219–224. doi:10.1016/j.conb.2005.03.006.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., Cappa, S.F., 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci. Biobehav. Rev. 35, 407–426. doi:10.1016/j.neubiorev.2010.04.007.

Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107, 829–853. doi:10.1093/brain/107.3.829.

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., van den Bosch, A., 2015. Prediction during natural language comprehension. Cereb. Cortex 26, 2506–2516. doi:10.1093/cercor/bhv075.

Wymbs, N.F., Bassett, D.S., Mucha, P.J., Porter, M.A., Grafton, S.T., 2012. Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. Neuron 74, 936–946. doi:10.1016/j.neuron.2012.03.038.

Yang, H., Lin, Q., Han, Z., Li, H., Song, L., Chen, L., He, Y., Bi, Y., 2017. Dissociable intrinsic functional networks support noun-object and verb-action processing. Brain Lang. 175, 29–41. doi:10.1016/j.bandl.2017.08.009.

Zhan, W., Guo, R., Chang, B., Chen, Y., Chen, L., 2019. The building of the CCL corpus: its design and implementation. Corpus Linguist. 6, 71–86.

Zhang, L., Pylkkanen, L., 2015. The interplay of composition and concept specificity in the left anterior temporal lobe: an MEG study. NeuroImage 111, 228–240. doi:10.1016/j.neuroimage.2015.02.028.

Zhang, L., Pylkkanen, L., 2018. Composing lexical versus functional adjectives: evidence for uniformity in the left temporal lobe. Psychon. Bull. Rev. 25, 2309–2322. doi:10.3758/s13423-018-1469-y.