Neural Tracking of Language Emerges from Distributed Synchronization, Sensitivity to Syntax, and Statistics

Filiz Tezcan

tezcanfiliz@gmail.com

Max Planck Institute for Psycholinguistics https://orcid.org/0000-0003-3327-0181

Cas Coopmans

Max Planck Institute for Psycholinguistics

Andrea Martin

Max Planck Institute for Psycholinguistics https://orcid.org/0000-0002-3395-7234

Article

Keywords:

Posted Date: January 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8269283/v1

Additional Declarations: There is **NO** Competing Interest.

Neural Tracking of Language Emerges from Distributed Synchronization, Sensitivity to Syntax, and Statistics

Abstract

Neural tracking, how brain activity synchronizes with speech, remains contentious regarding whether it's driven by lexical statistical information or linguistic structures. Previous research provided evidence for both accounts, but methodological limitations made direct comparisons difficult. We used isochronous Turkish stimuli with head-final phrase structure to eliminate confounds of word category repetition at the phrase level. By controlling acoustic features, word repetition, meaning-relatedness, transitional probabilities, and syntactic structure, we isolated distinct contributions of lexical statistics and syntactic structure to neural tracking. Results showed both factors independently shape neural responses, with complex linguistic structures engaging broader brain networks. Computational modeling revealed that hierarchically-coupled oscillations explained neural data better than statistical models alone, supporting structured linguistic representation building alongside word-level statistics. Our model proposes meaning emerges through flexible coupling among neural populations encoding lexical categories. These findings demonstrate that hierarchical processing crucially contributes to neural speech tracking beyond statistical repetitions.

Introduction

Neural tracking is the phenomenon by which the phase and power of neural signals become aligned with those of an external stimulus, often a sensory input, such as speech acoustics, but synchronization also occurs with the time course of perception of information that is endogenous or stems from stored knowledge, such as phonemes, morphemes, and words. Neural tracking has been argued to be a process underlying the temporal dynamics of speech perception that enables the brain to interpret such time-sensitive stimuli as language and music, which unfold over time. Importantly, neural tracking is not static but dynamically modulated by multiple linguistic dimensions, including semantic meaning, syntactic structure, word categories, word frequency, and statistical regularities in language (e.g., Di Liberto et al., 2015; Ding et al., 2016; Broderick et al., 2018; Kaufeld et al., 2020; Heilbron et al, 2020; Donhauser and Baillet, 2020; Coopmans et al., 2022, 2025; Ten Oever et al., 2022; Weissbart et al., 2020; Gillis et al., 2023; Tezcan et al., 2023, 2025; Mai and Wang, 2023; Slaats et al., 2023,2024; Karunathilake et al., 2024; Brodbeck et al., 2024; Zhao et al., 2024). Even as the number of empirical studies on neural tracking rapidly multiplies, what drives neural tracking still remains elusive (Meyer, Sun, and Martin, 2020; Ghitza, 2020; Giraud, 2020; Gwilliams, 2020; Haegens, 2020; Kandylaki & Kotz, 2020; Klimovich-Gray & Molinaro, 2020; Lewis, 2020). Central to this debate is the question of whether neural tracking is primarily mediated by surface-level properties, such as the statistical distribution of word-level information, or by more abstract processes like those which comprise the building and interpretation of hierarchical syntactic relations. For instance, the structure-building hypothesis posits that neural tracking reflects the hierarchical organization of linguistic input into progressively larger units, such as phrases and sentences, which recruits broader and more distributed networks in the brain (Marlsen-Wilson & Welch, 1978; Friederici, 2002; Hagoort, 2013, Pylkkänen, 2019; Martin 2016, 2020). This view is supported by studies demonstrating that neural entrainment occurs at multiple linguistic timescales, with slower oscillations (e.g., delta band) aligning with high-level sentence and phrase boundaries (Ding et al., 2016; Kösem et al., 2018; Kaufeld et al., 2020; Coopmans et al, 2022; Ten Oever et al., 2021) and faster oscillations (e.g., theta band) synchronizing with lower-level syllables and words (Pefkou et al., 2017; Mai and Wang, 2023)

In contrast, Frank and Yang (2018) suggested that neural tracking may reflect the brain's sensitivity to distributional properties in the input. Statistical regularities in speech, including patterns of word category co-occurrence and transitional probabilities, can shape neural responses without requiring any abstract representations. This view, we refer it as lexical-statistics hypothesis, is supported by evidence from distributional models, which can replicate neural tracking patterns observed in experimental data by relying solely on lexical relationships (Frank & Yang, 2018; Kalenkovich et al., 2022, see also Zhao et al., 2024).

The two views are difficult to disentangle, because they can account for the same patterns of results. Ding et al. (2016) conducted an influential study demonstrating that neural entrainment to speech reflects the neural tracking of hierarchical sentence structures. Using isochronously presented speech, they observed that only native speakers exhibit neural peaks at sentence and phrase frequencies, supporting the hierarchical processing in language comprehension. However, alternative interpretations challenge this view, suggesting that such neural entrainment may arise from surface-level lexical representations alone, without necessitating hierarchical structure-building. For example, in a sentence composed of a noun phrase and a verb phrase (e.g., Adj-Noun-Verb-Noun), nouns are repeated at intervals of 2 and 4 words. This repetition can generate peaks at phrase- and sentence-level frequencies based solely on recurring lexical patterns, such as word categories (Frank & Yang, 2018).

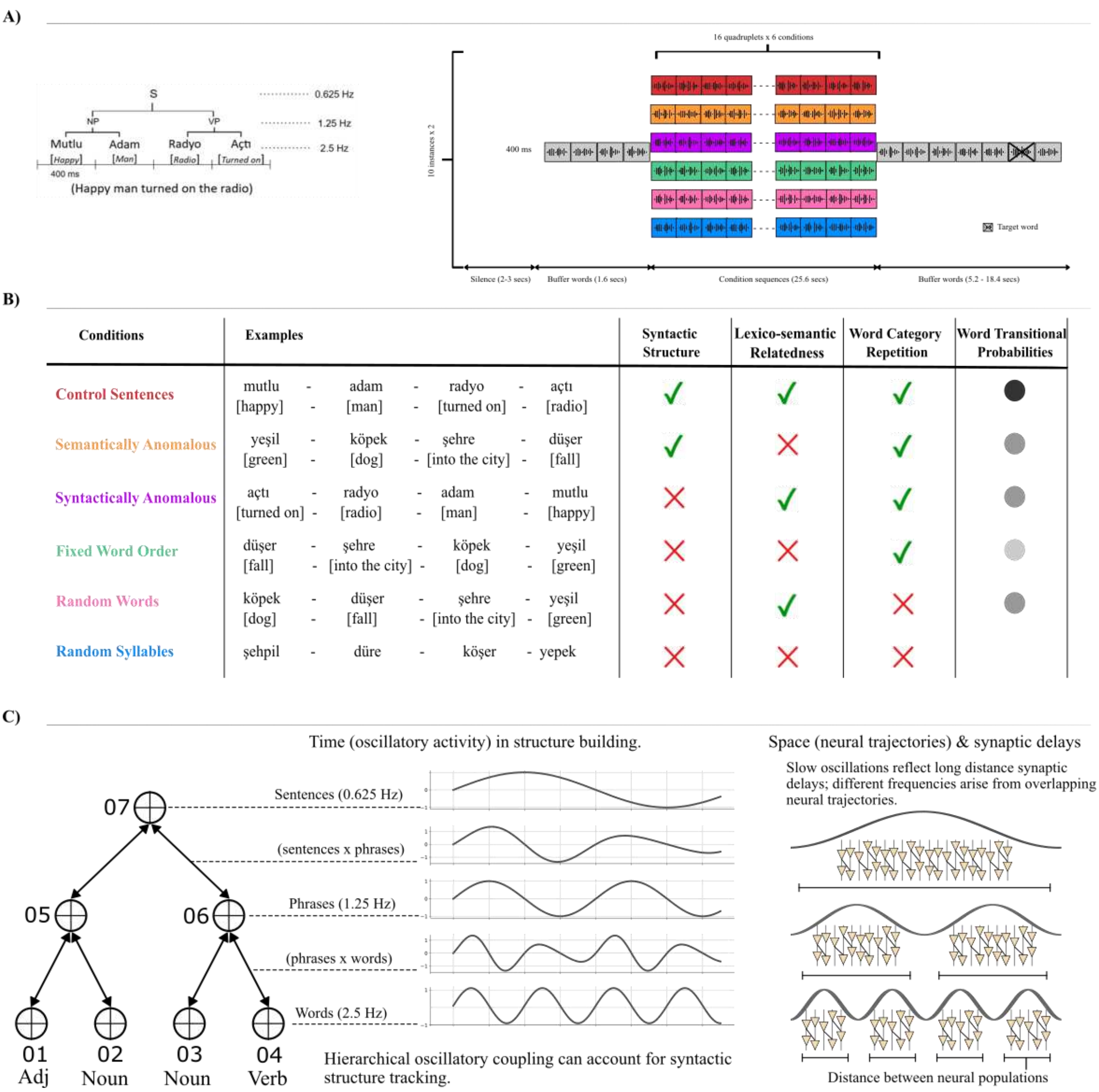
Efforts to disentangle these competing accounts have yielded mixed evidence. Studies have provided support for both hierarchical structure-building and lexical-statistics accounts (e.g., Burroughs et al., 2021; Lu et al., 2022; Lo et al., 2022; Zhao et al., 2024; Kalenkovich et al., 2022). However, methodological differences between these studies complicate direct comparisons. For example, some studies have relied on averaged power differences across all sensors (Lo et al., 2022; Lu et al., 2022; Burroughs et al., 2021; Zhao et al., 2024), while others have used stimuli with acoustically or lexically distinct conditions (Lu et al., 2023; Burroughs et al., 2021; Zhao et al., 2024; Kalenkovich et al., 2022), making it difficult to isolate the specific contributions of hierarchical structure versus lexical regularities.

In addition, very few studies have investigated the coordination of brain networks underlying neural speech tracking (e.g., Park et al., 2015; Schoffelen et al., 2017; Bai et al., 2022). Consequently, the precise ways in which connectivity patterns across distinct neural frequencies, corresponding to sentence, phrase, word, and syllable rates, are modulated by linguistic structure remain unclear. Further investigation is needed to understand how neural network coordination supports hierarchical and lexical processing during language comprehension.

To disentangle the contributions of lexical and syntactic information to neural speech tracking, this study employs a frequency-tagging paradigm using isochronous Turkish speech. The impact of varying levels of linguistic structure is examined through measures of induced power, inter-trial phase coherence (ITPC), and whole-brain functional connectivity. Additionally, we simulated the lexical model proposed by Frank and Yang (2018) and a novel hierarchically coupled oscillatory network model to provide further explanatory support for the observed neural data.

We propose that neural synchrony, operationalized through weakly coupled oscillators, facilitates the combination of neural representations through cross-frequency coupling (Lakatos et al, 2008; Lisman and Jensen, 2013; Arnal et al, 2015; Hyafil et al, 2015; Brennan and Martin, 2020; Ten Oever and Martin, 2021; Murphy, 2024; Weissbart and Martin, 2024). The network consists of three layers. The first layer includes four oscillators representing word categories (Adjective, Noun, Noun, Verb). The second layer contains two phrase nodes. The first phrase node is coupled to the Adjective and Noun oscillators, while the second phrase node is coupled to the Noun and Verb oscillators. In the third layer, a single sentence node is coupled to the phrase nodes from the second layer.

This architecture does not postulate dedicated oscillatory circuits for individual phrases or sentences. Instead, the model proposes that lexical categories are represented within low-dimensional latent spaces, encoded by overlapping neural populations (Ralph et al, 2017; Rogers et al., 2021) rather than requiring anatomical segregation as suggested by earlier neuroanatomical studies (Holmes et al., 1971; Damasio & Tranel, 1993; Shapiro et al., 2005). Neural population activity encoding word categories at distinct latent spaces according to semantic distributions may not function as a true (self-sustained fixed frequency) oscillator but rather as resonant activity with a preferred frequency. Resonant activity can amplify external stimuli near the population’s natural frequency and align its phase to even non-rhythmic inputs, providing flexible temporal gating for linguistic processing (Izhikevich et al, 2003; Doelling & Assaneo, 2021; Vinck et al., 2023).The compositional mechanism operates through cross-frequency coupling (Weissbart and Martin, 2024), whereby higher-order syntactic structures emerge from synchronization between low-dimensional resonant latent neural space of word categories (see Martin, 2016, 2020). For instance, adjective-noun phrasal composition is achieved through phase-locked synchronization between spatially overlapping but dynamically segregated neural ensembles representing adjective and noun categories, with the composite structure encoded by a slower oscillatory frequency in a wider neuronal population that binds the constituent elements. Phase-locked synchronization between oscillatory neural populations can be achieved via long range excitatory connections (See Figure 1 C).



Conditions	Examples	Syntactic Structure	Lexico-semantic Relatedness	Word Category Repetition	Word Transitional Probabilities
Control Sentences	mutlu - adam - radyo - açtı [happy] - [man] - [turned on] - [radio]	✓	✓	✓	●
Semantically Anomalous	yeşil - köpek - şehre - düşer [green] - [dog] - [into the city] - [fall]	✓	✗	✓	●
Syntactically Anomalous	açtı - radyo - adam - mutlu [turned on] - [radio] - [man] - [happy]	✗	✓	✓	●
Fixed Word Order	düşer - şehre - köpek - yeşil [fall] - [into the city] - [dog] - [green]	✗	✗	✓	●
Random Words	köpek - düşer - şehre - yeşil [dog] - [fall] - [into the city] - [green]	✗	✓	✗	●
Random Syllables	şehpil - düre - köşer - yepek	✗	✗	✗	

Figure 1 A. Experimental Design B. Stimuli C. Hierarchically coupled oscillatory network model

Our stimuli were in Turkish, which is well-suited for our purposes. Turkish phrases are head-final, which means that the verb follows its object (*kurabiyeleri yedi*, lit. "cookies ate"), in contrast to the head-first order in English ("ate cookies"). This feature of Turkish is useful because it avoids any repetition of word categories at the phrase rate in regular sentences when it follows a noun phrase such as *aç kadın kurabiyeleri yedi* "[*hungry*] [*woman*] [*cookies*] [*ate*]" (Adjective-Noun-Noun-Verb)

We constructed speech streams in six experimental conditions designed to systematically vary linguistic structure while maintaining control over acoustic properties (Figure 1 B). Each stream consisted of 64 disyllabic words, which formed 16 four-word sequences. We used the same syllables in each condition organized in different orders to minimize acoustic variability across conditions. The conditions contain progressively more linguistic information. In Random Syllable Lists, consecutive syllables could not be combined to form real disyllabic words. This resulted in sequences devoid of linguistic structure at the word, phrase, or sentence levels. In Random Word Lists, two consecutive syllables formed a word, but two consecutive words did not combine into phrases or sentences. However, words within each quadruplet were semantically related and could form a meaningful sentence if reordered correctly. In Fixed Word Lists, consecutive words did not form phrases or sentences, but word categories (e.g., Verb-Noun/Adv-Noun-Adj) were repeated every four words. Words within each quadruplet were not semantically related. Syntactically Anomalous Sentences, are similar to Fixed Word Lists in the sense that word categories were repeated every four words. However, the words within each quadruplet were semantically related and could form a meaningful sentence if reordered correctly. In Semantically Anomalous Sentences, quadruplets were structured such that the first two words formed an anomalous noun phrase and the last two an anomalous verb phrase (e.g., Adj-Noun-Noun/Adv-Verb). This created sentences that were syntactically coherent but semantically nonsensical. Last, Control

Sentences were composed of quadruplets that formed fully coherent sentences, both syntactically and semantically. In sum, constructing conditions that have exactly the same syllables in different orders allows us to systematically vary linguistic structure while controlling for acoustic and lexical properties, thus enabling a targeted examination of language processing under different structural constraints.

Moreover, specific comparisons allow us to dissect the neural signatures associated with all levels of linguistic organization, offering insights into the processing of lexical, syntactic, and semantic elements in sentence comprehension. **Random Word List vs. Random Syllables**: This contrast was used to investigate the impact of word structure on neural responses. **Fixed Word Order vs. Random Syllables**: By comparing these conditions, we assessed the combined effects of word structure and the regular repetition of word categories, which introduces a predictable organizational pattern without sentence level syntactic structure. **Control Sentences vs. Semantically Anomalous**: This comparison examined the influence of semantic coherence and/or transitional probabilities within sentences, highlighting how meaning and statistical patterns modulate neural processing. **Control Sentences vs. Syntactically Anomalous**: This contrast primarily reflected the effects of syntactic structure. However, as the syntactic manipulation also involves differences in statistical patterns, the findings should be interpreted with this overlap in mind**. Semantically Anomalous vs. Syntactically Anomalous:** This comparison isolates the neural contributions of syntactic structure by controlling for transitional probabilities between consecutive words, as both conditions consist of low-transitional probability sequences. In the Semantically Anomalous condition, syntactic structure is preserved, but transitional probabilities are low due to phrase combinations that violate lexical knowledge. Conversely, in the Syntactically Anomalous condition, low transitional probabilities arise from violations of syntactic structure. By equating statistical differences between conditions, this comparison eliminates transitional probability as a confounding factor, allowing for the specific investigation of syntactic processing. **Control Sentences vs. Random Syllables**: This final contrast captured the cumulative effects of word structure, semantic coherence (and/or transitional probabilities between consecutive words), and syntactic organization within sentences.

Results

Behavioral Results

To assess whether participants paid attention to the stimuli, we analyzed their performance on multiple-choice questions by comparing the percentage of correct responses. Behavioral data from two participants were unavailable due to a mistake by the experimenter so the analysis included data from 28 participants. On average, participants answered 97.6% (SD = 4.59%) of the questions correctly, which was significantly above the chance level of 25% ($t(27)=110.3$, $p<0.0001$).

Power Spectra of the Stimuli

Analyzing the power spectra of the stimuli, we found that all conditions exhibited a significant spectral peak at both the word (2.5 Hz) and syllable (5 Hz) frequencies ($p < 0.05$), while no significant peaks were observed at the sentence (0.625 Hz) or phrase (1.25 Hz) frequencies. (Figure 1A). Comparison of powers at each frequency did not show any significant effect of condition (Sentence Freq. $F(5,54) = 1.461$, $p=0.218$; Phrase Freq. $F(5,54) = 0.062$, $p=0.997$; Word Freq. $F(5,54) = 0.324$, $p=0.897$; Syllable Freq. $F(5,54) = 0.084$, $p=0.994$ See Figure 1 and Suppl. Figure 1)

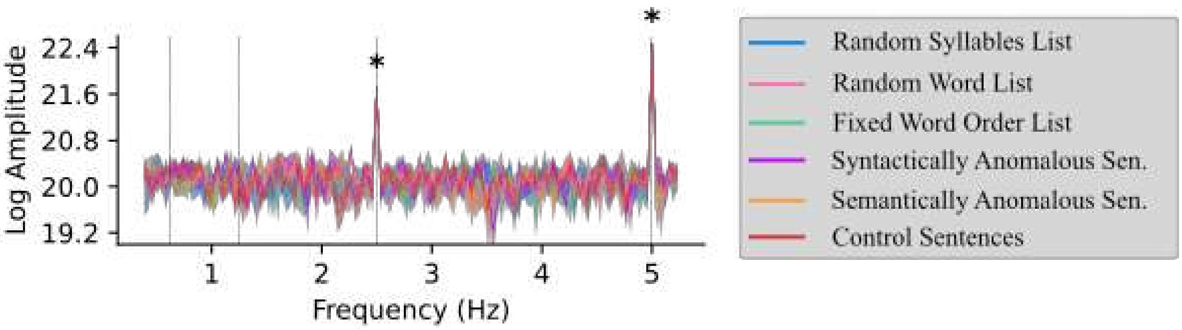


Figure 1. A) Power spectrum of stimuli for each condition (Vertical dashed lines indicate the FOIs at sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz), and syllable frequency (5 Hz).

Stars indicate significant peaks ($p<0.05$); shaded areas indicate the standard error of the mean over trials.)

The cosine similarity of word pairs within each quadruplet, used as a measure of lexico-semantic relatedness, was significantly higher in the Control Sentences, Syntactically Anomalous, and Random Words List conditions compared to the Semantically Anomalous and Fixed Word Order List conditions ($t(1) = 2.65$, $p = 0.0079$). Additionally, analysis of the surprisal values for words across trials revealed a significant effect of condition ($F(4,3195) = 35.99$, $p < 0.0001$). Pairwise comparisons indicated that surprisal values were highest in the Fixed Word Order List condition and lowest in the Control Sentences condition relative to all other conditions. No significant differences in surprisal values were observed between the Random Words List, Semantically Anomalous, and Syntactically Anomalous conditions (see Supplementary Fig. 2 and Table 1). These results confirm that the experimental conditions were effectively designed to manipulate the semantic relatedness of words within quadruplets.

Power Spectra of the MEG Signal

All conditions exhibited a significant peak ($p < 0.05$) at the word (2.5 Hz) and syllable (5 Hz) frequencies. The Control Sentences, Semantically Anomalous Sentences, and Fixed Word List conditions additionally showed significant peaks at both sentence (0.625 Hz) and phrase (1.25 Hz) frequencies, along with significant peaks at the harmonics of the sentence frequency. While the peak at the sentence frequency was not significant in the Syntactically Anomalous Sentences condition, the phrase frequency peak was significant, and significant peaks were also observed at the harmonics of the sentence frequency (Figure 2A).

LMM results elicited significantly different spectral power for conditions at all frequencies (Sentence Frequency ($F(5,145) = 6.29$, $p < 0.0001$), Phrase Frequency ($F(5,145) = 23.6$, $p <$

0.0001), Word Frequency ($F(5,145) = 69.79$, $p < 0.0001$), and Syllable Frequency ($F(5,145) = 5.07$, $p < 0.001$).

At the sentence frequency, the pairwise comparison analysis revealed that the power in the Control Sentences condition was significantly greater than that of the Syntactically Anomalous, Random Words List, and Random Syllables conditions (see Figure 2B and Suppl. Table 2).

At the phrase frequency, the power in the Control Sentences condition remained significantly higher than all other conditions, with the exception of the Semantically Anomalous condition. Notably, although the power in the Semantically Anomalous condition at the sentence frequency did not significantly differ from that in the Syntactically Anomalous condition, it was greater than that at the phrase frequency level.

At the word frequency, the Random Syllables condition showed significantly lower power compared to all other conditions, despite being significantly not different than other conditions in the stimuli.

At the syllable frequency level, the power in the Random Syllables condition was significantly greater than all other conditions, even though these conditions did not differ acoustically at this frequency. Figure 2B illustrates the power spectra across all conditions at each frequency of interest (FOI).

We further examined source-level power differences across conditions by identifying brain regions where significant power differences occurred at each frequency, using a permutation cluster t-test. Power differences were predominantly distributed across the bilateral language network for each contrast. However, we observed a reduced power at the syllable frequency in the contrasts between Control Sentences and Random Syllable List, as well as Random

Word List and Random Syllable List, and it was localized on the right hemisphere (See Figure 2C).

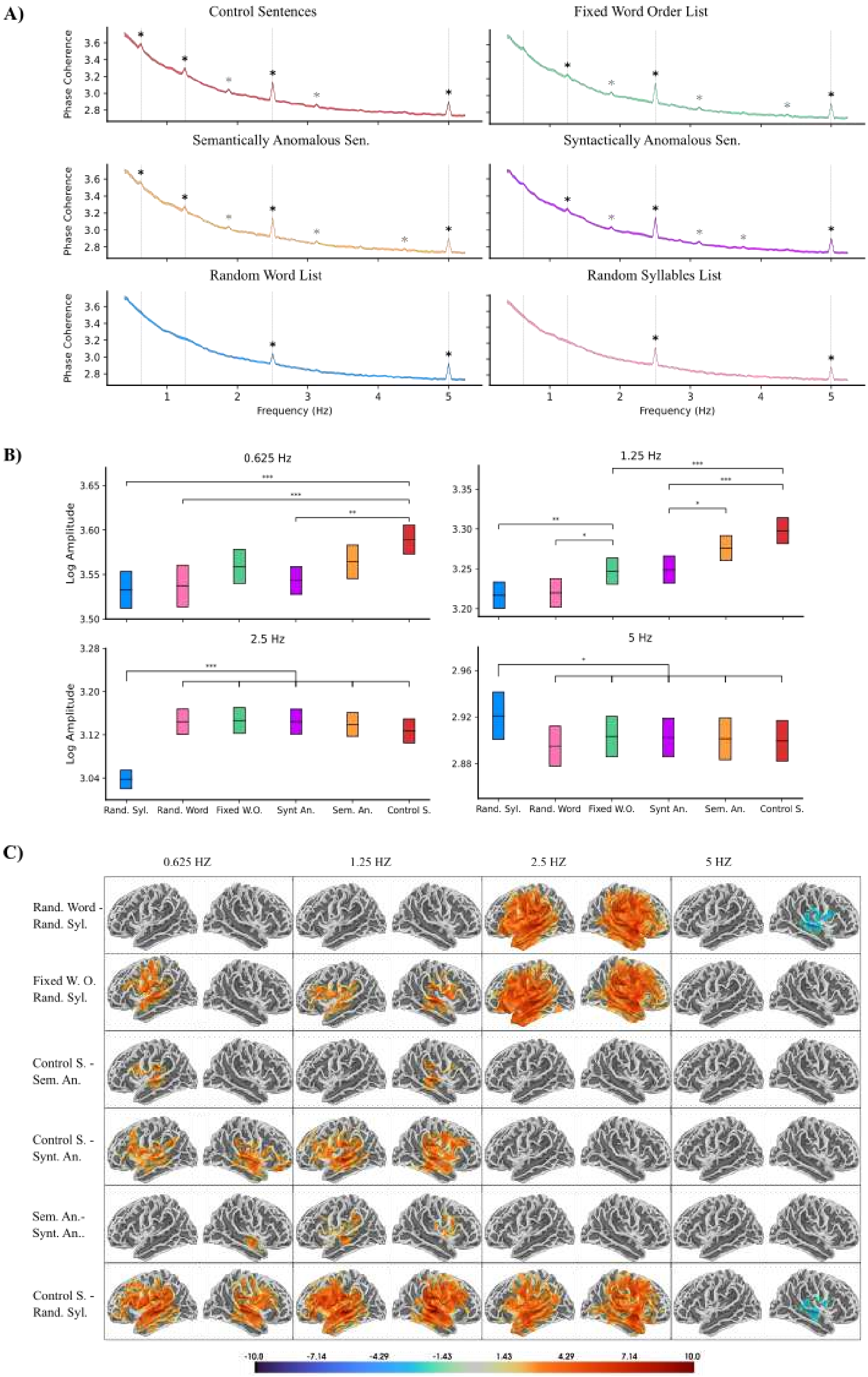


Figure 2. A) Power spectrum of whole brain averaged neural signal in each condition. The vertical dashed lines indicate the FOIs at the sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz), and syllable frequency (5 Hz), and black asterisks indicate significant peaks at FOIs, and the grey asterisks indicate significant peaks at the harmonic frequencies (**** <0.0001, *** <0.001, **<0.01, * < 0.05). Shaded areas indicate the standard error of the mean over subjects **B)** Comparison of power averaged over the whole brain at FOIs. Box edges indicate the standard error of the mean over subjects. Black line shows the mean. **C)** Sources that were significantly different between the contrasts. The colors indicate the t-values of the corresponding comparisons.

Intertrial Phase Coherence (ITPC) of the MEG Signal

ITPC analysis showed similar results to the induced power comparison except that ITPC at sentence frequency (0.625 Hz) in Syntactically Anomalous Sentences condition also showed a significant peak (Figure 3A).

LMM also revealed significantly different ITPC values for conditions at all frequencies (Sentence Frequency: $F(5,145) = 75.51$, $p < 0.0001$, Phrase Frequency: $F(5,145) = 80.39$, $p < 0.0001$, Word Frequency: $F(5,145) = 63.07$, $p < 0.0001$, and Syllable Frequency: $F(5,145) = 6.29$, $p < 0.0001$; see Figure 3 B, Suppl Table 3)

Similar to the induced power comparison, at the sentence frequency, the pairwise comparison analysis indicated that ITPC in the Control Sentences condition was significantly higher than in all other conditions. ITPC in the Semantically Anomalous condition was also significantly higher than in the Syntactically Anomalous, Random Word List, and Random Syllable Lists conditions but did not differ significantly from the Fixed Word Order condition.

At the phrase frequency level, ITPC in the Control Sentences condition remained significantly greater than in all other conditions. Although ITPC in the Semantically

Anomalous condition at the sentence frequency was not significantly different from the Fixed Word Order condition, it exceeded ITPC in the Fixed Word Order condition at the phrase frequency level.

At the word frequency level, ITPC in the Random Syllables condition was significantly lower than in all other conditions. Conversely, at the syllable frequency level, ITPC in the Random Syllable List condition was significantly higher than in all other conditions.

Further examination of source-level ITPC differences across conditions, also revealed similar results to induced power comparison, with the exception that in the ITPC analysis, no significant difference was observed at the syllable frequency between Control Sentences and Random Syllables or between Random Word List and Random Syllable List (see Figure 3C).

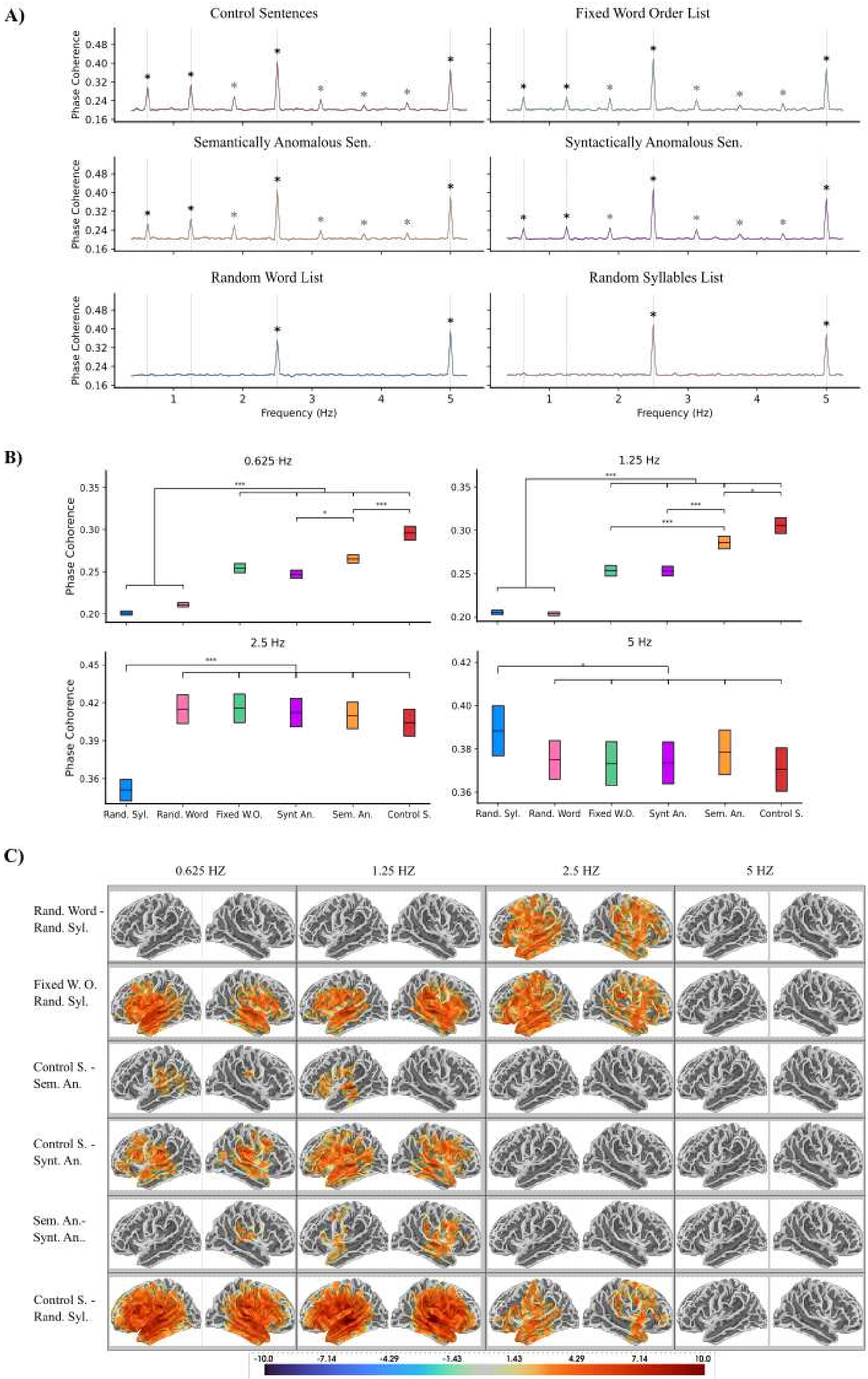


Figure 3. A) Whole brain averaged ITPC of neural signal in each condition (Vertical dashed lines indicate the FOIs at the sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz) , and

syllable frequency (5 Hz); black stars indicate significant peaks at FOIs and grey stars indicate significant peaks at the harmonic frequencies) (**** <0.0001, *** <0.001, **<0.01, * < 0.05). Shaded areas indicate the standard error of the mean over subjects **B)** Comparison of ITPC over whole brain at FOIs. Box edges indicate the standard error of the mean over subjects. Black line shows the mean. **C)** t values of the contrasts on the source level (color bar indicate the t values at significant sources)

Functional Connectivity Analysis

Figure 4 shows significant differences between conditions in terms of functional connectivity (pairwise phase coherence) between brain regions. It can be seen that neural synchronization between different brain regions intensifies as syllable streams are organized into more complete linguistic structures, reflecting the hierarchical processing of language. When syllables form real words, evidenced by the comparison between the *Random Word Lists* and *Random Syllable Lists*, functional connectivity increases between the right parietal and frontal cortices, as well as within the left temporal cortex, particularly at the word frequency (2.5 Hz) .

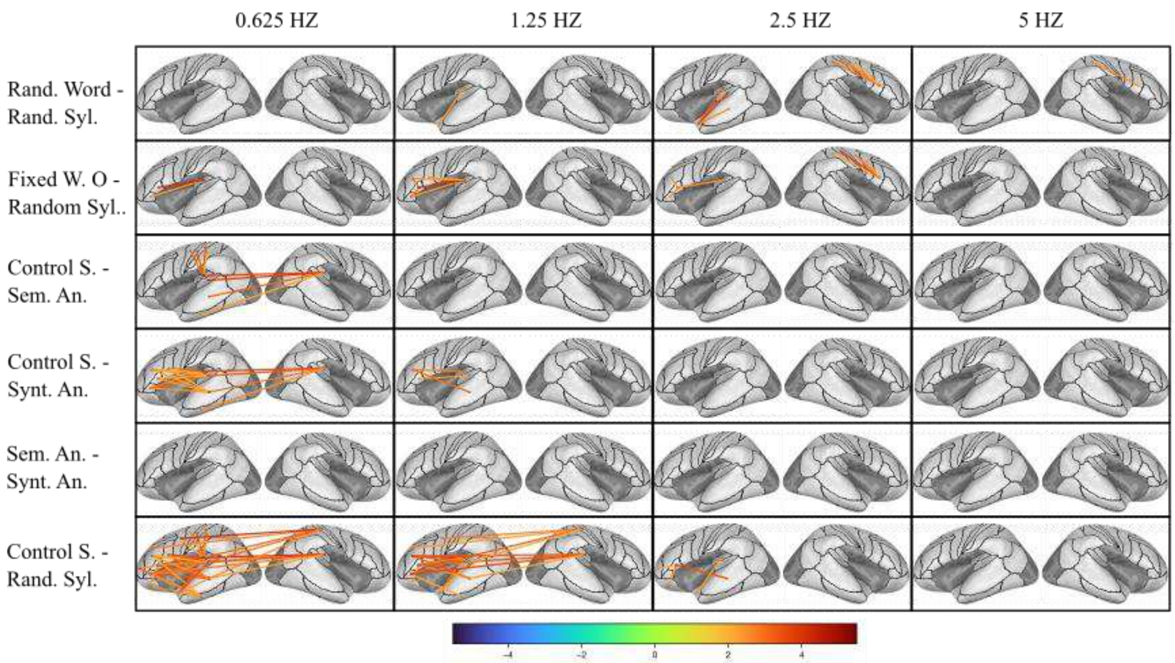


Figure 4. t values of connectivity comparison between conditions (color bar indicate the t values at significant regions)

The organization of four consecutive words can exhibit regular patterns based on semantic relationships, syntactic structures, or categorical classifications. In the *Fixed Word Order* condition, where word categories are repeated, connectivity is enhanced between the auditory and frontal cortices in the left hemisphere across sentence, phrase, and word frequencies compared to the *Random Syllables* condition. Additionally, this condition also shows increased connectivity between the parietal and frontal cortices in the right hemisphere, similar to the connectivity pattern observed in the word-level contrast.

Semantic coherence within quadruplets, examined through the contrast between *Control Sentences* and *Semantically Anomalous Sentences*, further increases neural synchrony between the right and left temporal cortices, as well as between the left auditory and motor cortices.

Syntactic structure, assessed through the contrast between *Control Sentences* and *Syntactically Anomalous Sentences,* like the semantic coherence, also increases functional connectivity between the right and left temporal cortices, and in addition to that there is an increase in functional connectivity between the left temporal and frontal cortices.

The contrast between Semantically and Syntactically Anomalous conditions which has different words in each quadruplet didn't show any significant connectivity difference.

Finally, the contrast between the most structured (*Control Sentences*) and least structured (*Random Syllables*) conditions reveals the most widespread increase in functional connectivity.

In summary, increasing linguistic structure is associated with enhanced connectivity between brain regions at slower frequency bands, engaging a broader neural network. Connectivity driven by word-level structure predominantly occurs at the word frequency, while sentence-level structure elicits increased connectivity at both the phrase and sentence frequencies.

Hierarchically Coupled Oscillatory Network Model

We simulated the neural response to all conditions except the Random Syllable List condition because the oscillatory model doesn't have syllable level nodes. All conditions demonstrated significant peaks ($p < 0.05$) at the word frequency (2.5 Hz). Additionally, the Fixed Word, Syntactically Anomalous, Semantically Anomalous, and Control Sentences conditions exhibited significant peaks at the sentence frequency (0.625 Hz) and phrase frequency (1.25 Hz). Analysis of variance (ANOVA) revealed significant differences in spectral power across experimental conditions at both the sentence- and phrase-level frequencies, whereas no significant differences were observed at the word- or syllable-level frequencies (Sentence Frequency ($F(4,145) = 265.9$, $p < 0.0001$), Phrase Frequency ($F(4,145) = 294.3$, $p < 0.0001$), Word Frequency ($F(4,145) = 0.46$, $p=0.90$), and Syllable Frequency ($F(4,145) = 0.81$, $p =0.39$)). Pairwise comparisons for each frequency band power are shown on Figure 6 B (see Suppl. Table 1).

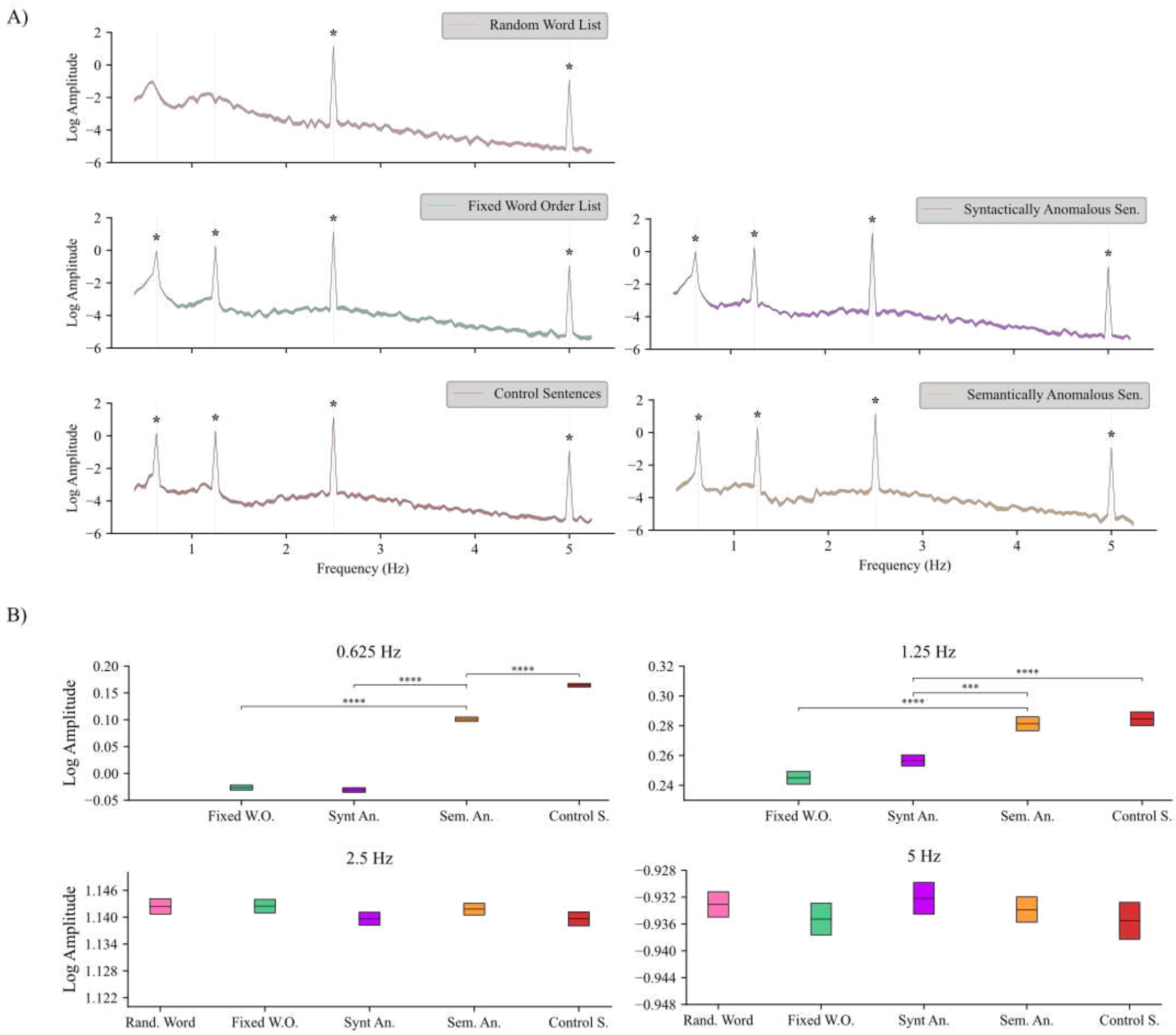


Figure 5. A) Power spectrum of simulated conditions (Vertical lines mark sentence, phrase, word and syllable frequencies, * indicates significant peaks) **B)** Power at sentence-, phrase-, and word-level frequencies (**** <0.0001, *** <0.001, ** <0.01, * < 0.05)

Lexical Model

We simulated the neural response to all conditions except the Random Syllable List condition because it does not contain real words and therefore could not be modeled via word embeddings. All conditions demonstrated significant peaks ($p < 0.05$) at the word frequency (2.5 Hz) and syllable frequency (5 Hz). Additionally, the Fixed Word, Syntactically Anomalous, Semantically Anomalous, and Control Sentences conditions exhibited peaks at the sentence frequency (0.625 Hz) and phrase frequency (1.25 Hz). However, the phrase-level peak in the Control Sentences condition did not reach statistical significance. Analysis of variance (ANOVA) of the lexical model at each FOI revealed a significant effect of condition at sentence and phrase frequencies (Sentence: $F(4,45) = 60.01$, $p<0.0001$; Phrase: $F(4,45) = 20.22$, $p<0.0001$; Word: $F(4,45) = 0.668$, $p=0.618$; Syllable: $F(4,45) = 0.737$, $p=0.571$). Pairwise comparison showed that at both sentence and phrase frequency, power of Fixed Word, Syntactically Anomalous, Semantically Anomalous and Control Sentences conditions is higher than Random Words condition. In contrast to the neural data, the simulated data did not show a difference between the Control Sentences and Fixed Word Lists, Syntactically Anomalous and Semantically Anomalous Sentences, neither either at the phrase rate nor at the sentence rate (Figure 6, Suppl. Table 4).

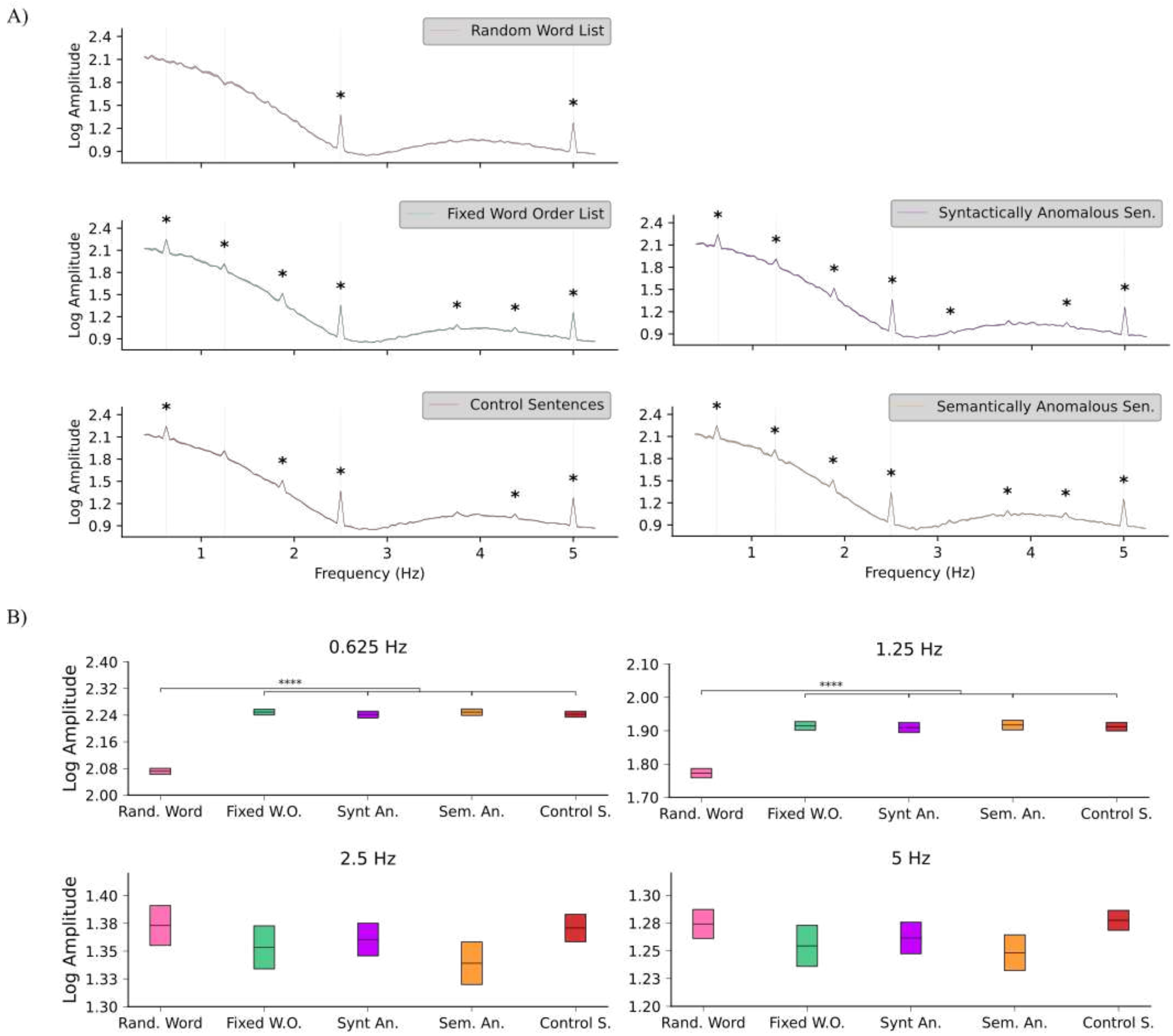


Figure 6. A) Power spectrum of lexical model in each condition. The vertical dashed lines indicate the FOIs at the sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz), and syllable frequency (5 Hz), and black asterisks indicate significant peaks at FOIs, and the grey asterisks indicate significant peaks at the harmonic frequencies (**** <0.0001, *** <0.001, **<0.01, * < 0.05). Shaded areas show the standard error of the mean within trials **B)** Comparison of power averaged over trials. Box edges indicate the standard error of the mean over trials. Black line shows the mean.

Discussion

We utilized a frequency-tagging paradigm with isochronous speech to examine how the presence of varying levels of linguistic structure influenced neural speech tracking. The primary goal of this study was to disentangle the distinct contributions of syntactic structure, semantic coherence (and/or transitional probabilities between consecutive words), and word category repetition. Specifically, we aimed to test two competing hypotheses regarding the processes underlying neural speech tracking: (1) the Lexical-Statistic Hypothesis, which posits that repetition of lexical-level distributional representations can generate power peaks at phrase- and sentence-level frequencies, and (2) the Structure-Building Hypothesis, which proposes that lexical information is integrated into hierarchical structure.

To address these hypotheses, we analyzed measures of induced power, inter-trial phase coherence, and whole-brain functional connectivity. Additionally, we simulated a computational model representing the Lexical-Statistic Hypothesis.

Our findings revealed that neural tracking increases at word, phrase, and sentence frequencies as syllables coalesce into words, words combine into phrases, and phrases combine into sentences. The results of connectivity analyses suggested that this increase in neural tracking is driven by the synchronization of an increasingly extensive neural network as the complexity of linguistic representations rises. Neural activity within lower-frequency bands exhibits greater synchronization across a wider brain network. Furthermore, semantic coherence and word category repetition also contributed to enhanced neural tracking, alongside syntactic structure, although these factors appeared to engage distinct neural networks.

Additionally, our computational simulations demonstrated that a non-hierarchical lexical model, in which distributional vectors are repeated, can simulate neural tracking at sentence and phrase frequencies in condition which word categories are repeated. However, this model did not account for contributions of syntactic structure. In contrast, a hierarchically coupled oscillatory network, where word-category nodes are coupled with a phrase node, and phrase nodes are coupled with a sentence node, can successfully distinguish between semantic and syntactic coherence. This model does not assume fixed oscillators or dedicated circuits for individual words or phrases. Instead, it generates compositional structure dynamically through cross-frequency coupling within low-dimensional resonant activity, supported by overlapping neural populations that encode lexical categories.

In the following sections, we will discuss each of these findings in order of increasing linguistic complexity.

When syllables combine into words, neural tracking at word frequency was enhanced and it was suppressed at the syllable frequency:

We analyzed induced power to examine the modulation of power that is not necessarily phase-aligned and inter-trial phase coherence (ITPC) to separately investigate phase alignment. Whole-brain average results from the Random Words versus Random Syllables contrast revealed that both power and ITPC at the word frequency increased when syllables formed words, while neural tracking at the syllable frequency was stronger in the Random

Syllables condition. At the source level, we observed increases in both induced power and phase coherence at the word frequency in widely distributed cortical regions for words. However, when comparing induced power in Random Words versus Random Syllables at the syllable frequency, we identified suppression around the right auditory cortex for words or enhancement for syllables. This modulation may reflect reduced reliance on lower-level information due to higher-level expectations for the second syllable in real words. In contrast, in the Random Syllables condition, the first syllable of a real word is followed by the second syllable of a different word so it was unexpected (Rao & Ballard, 1999; Friston, 2005; Sohoglu et al., 2024). Alternatively, this effect may indicate that linguistic structure enhances the tracking of higher-level representations (Kaufeld et al., 2020; Coopmans et al., 2022; Ten Oever et al., 2022; Tezcan et al., 2023, 2025; Weissbart and Martin, 2024; Mai & Wang, 2023; Gillis et al., 2023; Brodbeck et al., 2024; Pérez-Navarro, 2024). In the Random Words condition, the highest level of linguistic information processed corresponds to words, whereas in the Random Syllables condition, it is limited to individual syllables. However, since this contrast did not control for statistical regularities at the syllable or sub-word level, nor for the structural properties of word formation, it remains unclear whether the observed suppression or enhancement at word and syllable frequencies is driven by statistical properties or hierarchical linguistic structure. Consequently, the present experimental design does not allow for a definitive dissociation between these potential processes.

Word category repetition increased neural tracking at both sentence and phrase frequency:

The contrast between Fixed Word List and Random Syllables revealed a similar pattern of word formation at the word frequency. Additionally, it demonstrated an increased power and phase coherence at both phrase and sentence frequencies. Notably, since word categories were only repeated every four words (sentence frequency), and since there were no phrases in the Fixed Word List condition, there was no regular repetition at the phrase frequency. This suggests that the increase at the phrase frequency could potentially represent a harmonic of the sentence frequency (Zhou et al, 2016). These findings align with the lexical-statistics hypothesis (Saffran et al, 1996; Frank et al., 2012; Batterink and Paller, 2017, 2019; Frank and Christiansen, 2018; Frank and Yang, 2018) suggesting that repetition of word categories can generate a power peak at the sentence frequency. Furthermore, tracking the repetition of word categories enhanced phase synchronization between the left auditory and frontal cortices.

In this comparison, the suppression of induced power at the syllable frequency for words which was observed in the previous contrast was not evident for source level comparison of the Fixed Word List condition. However, whole-brain-averaged power at the syllable frequency remained higher in the Random Syllables condition compared to the Fixed Word List condition. If this suppression were driven by predictive processing related to the expectation of a second syllable in real words or tracking the highest-level linguistic structure, we would expect the same effect to emerge in this comparison, as well as in the comparison of Random Syllables with all other conditions, since all conditions used the same set of words. To further investigate, we directly compared Random Syllables with all other conditions. This analysis revealed suppressed induced power at the syllable frequency in all other conditions localized around the right auditory cortex (see Supplementary Figure 3). Notably, the statistical significance of the cluster exhibiting reduced induced power for words in this contrast approached conventional thresholds, with a p-value of 0.071.

Semantic coherence in a sentence (and/or transitional probabilities between consecutive words) also contributed to the neural tracking:

The contrast between Control Sentences and Semantically Anomalous Sentences highlighted the contribution of semantic coherence (and/or transitional probabilities between words) within a sentence. This was associated with increased induced power and phase coherence at both sentence and phrase frequencies, as observed in the source-space analysis. However, the effect was more pronounced in phase coherence, as demonstrated in the whole-brain average analysis as well. Since both conditions share the same syntactic structure and pattern of word category repetition, these differences can be attributed to the presence or absence of semantic coherence (Kaufeld et al., 2020; Coopmans et al., 2022) and/or the statistical regularities (transitional probabilities between consecutive words) informed by broader language knowledge. This process appears to enhance phase synchronization between the right auditory and left temporal cortices, as well as between the left auditory and parietal cortices.

Syntactic structure enhanced neural tracking beyond statistical information alone:

To examine the effect of syntactic structure on neural tracking, we compared Control Sentences with Syntactically Anomalous Sentences. In the Syntactically Anomalous condition, word order was reversed, thereby preserving word category repetition and the presence of semantically related words within each quadruplet (see also Lo et al., 2022). However, this manipulation introduced violations of transitional probabilities between consecutive words as well as syntactic structure (Slaats and Martin, 2025). The combination of syntactic structure and transitional probabilities enhanced induced power and inter-trial phase coherence (ITPC) at sentence- and phrase-level frequencies. Similar to the comparison between Control Sentences and Semantically Anomalous Sentences which showed the effect of semantic coherence and/or transitional probabilities, this contrast also revealed increased functional connectivity between the right and left temporal cortices. Additionally, there was an observed increase in functional connectivity between the left temporal and frontal cortices, likely driven by syntactic processing (Hagoort, 2005; Friederici, 2011)

To isolate the specific contribution of syntax, we compared Semantically Anomalous Sentences with Syntactically Anomalous Sentences. Both conditions exhibited similar surprisal values (i.e., transitional probabilities) between consecutive words and had repeated word categories every four words, but they lacked shared lexical items within each quadruplet. The Semantically Anomalous Sentences compared to Syntactically Anomalous sentences demonstrated increased induced power and ITPC at frequencies associated with sentence- and phrase-level structures, indicating that syntactic structure enhances neural tracking beyond what can be attributed to word category repetition and transitional probabilities (Ding et al 2016; Zhao et al. 2024; Coopmans et al. 2025; Weissbart and Martin, 2024). However, no significant differences in functional connectivity were observed between these conditions. This absence of connectivity differences may reflect the engagement of distinct cognitive strategies for integrating information within word quadruplets. In the Semantically Anomalous Sentences, syntactic structure drives the integration process, whereas in the Syntactically Anomalous Sentences, semantic relatedness of words likely played a more dominant role. These divergent mechanisms may recruit different neural pathways which may not be distinguished phase coherence analysis such as the direction of connectivity, contributing to the observed sentence-level tracking in both conditions.

The Lexical Model accounted for sentence- and phrase-level peaks through word category repetition but did not capture syntactic structure, a limitation overcame by the Oscillatory Model

We simulated neural responses using the Lexical Model proposed by Frank and Yang (2018). As consistent with previous studies (Burroughs et al., 2021; Lo et al., 2022; Kalenkovich et al., 2022) our results revealed that the Lexical Model produces peaks at phrase- and sentence-level frequencies across all conditions involving word category repetition. However, the model failed to account for the effects of syntactic structure on tracking as evidenced by statistically similar peak magnitudes observed for Control and Syntactically Anomalous sentences, as well as for Semantically Anomalous and Fixed Word Order conditions. This would be consistent with embedding vectors generated by the GPT-2 model encoding lower-dimensional patterns that to some degree may reflect word categories. But these lower-dimensional patterns do not capture transitional probabilities between word categories, as the transitional probabilities between words were primarily captured at the token level rather than generalized to the level of word category.

Furthermore, our analysis of surprisal values revealed lower values for Control Sentences compared to Semantically Anomalous and Syntactically Anomalous sentences, with no significant difference between the latter two conditions. Importantly, this lack of differentiation persisted despite the intact grammatical structure of Semantically Anomalous sentences. These could be one of the reasons that while models like GPT-2 excel in generating coherent and predictable sentences and continuations, they struggle to generalize effectively to novel combinations of word categories due to their reliance on token-based relationships rather than abstract categories and hierarchical relationships between them.

The Hierarchically Coupled Oscillatory Network, similar to DORA (Martin & Doumas, 2017), integrates word categories into higher-order abstract nodes, such as phrases and sentences. We propose that neural synchrony, operationalized through coupled oscillators, facilitates the combination of neural representations through cross-frequency coupling (Lakatos et al, 2008; Lisman and Jensen, 2013; Arnal et al, 2015; Hyafil et al, 2015; Brennan and Martin, 2020; Ten Oever and Martin, 2021; Murphy, 2024; Weissbart and Martin, 2024). This approach aligns with the notion that linguistic unit representations, activated by sensory input, must be maintained in a separable and active state until their transformation into higher-order constructs is complete (Poeppel & Monahan, 2011; Bever & Poeppel, 2010; Marslen-Wilson & Welsh, 1978; Halle & Stevens, 1962; Martin, 2020).

Our findings demonstrate that coupled oscillators associated with word category nodes can account for differences in power magnitude at phrase- and sentence-level frequencies across conditions. These oscillators generate distinct neural responses that reflect changes in both syntactic structure and semantic coherence and/or transitional probabilities, thereby capturing the hierarchical structure of linguistic input.

In conclusion, we investigated how varying levels of linguistic structure influence neural speech tracking using a frequency-tagging paradigm with isochronous speech. By examining induced power, inter-trial phase coherence, and whole-brain functional connectivity, we aimed to disentangle the contributions of syntactic structure, semantic coherence, and word category repetition. Our results demonstrated that neural tracking is progressively enhanced at word, phrase, and sentence frequencies as linguistic elements are hierarchically integrated into higher-order structures. Word category repetition, semantic coherence (and/or transitional probabilities), and syntactic structure each independently contributed to enhanced neural tracking. Complementary connectivity indicated that this enhancement is associated with the recruitment and synchronization of an increasingly distributed neural network reflecting the growing complexity of linguistic processing.

To further explore the underlying computational principles, we simulated a model suggested by Frank and Yang (2018). A non-hierarchical lexical model, in which lexical distributions are repeated, successfully replicated power peaks at phrase- and sentence-level frequencies for conditions which word categories are repeated but failed to account for syntactic contributions. In contrast, a hierarchically coupled oscillatory model, where word-category nodes interact with phrase and sentence nodes, effectively captured the distinction between syntactic structure and semantic coherence. These findings support the idea that hierarchical processing plays a crucial role in neural speech tracking beyond simple word-category repetition. They suggest that cross-frequency–coupled oscillatory activity may help organize hierarchical structure, and both statistical information and syntactic structure contribute to shaping linguistic representations in the brain.

Materials and Methods

Participants

Thirty native speakers of Turkish participated in the experiment (17 females, mean age = 32.4 years, age range = 22-42 years). All participants were right-handed, reported normal hearing, had either normal or corrected-to-normal vision, and had no history of dyslexia or other language-related disorders. Prior to participating in the MEG and MRI sessions, participants underwent a screening for eligibility and provided written informed consent. The study received approval from the Ethics Committee for Human Research in Arnhem/Nijmegen (project number CMO2014/288). Participants were compensated for their involvement in the study.

Stimuli

The experiment consisted of six conditions (see Table 1 for examples). Each condition had 10 unique syllable streams, each repeated twice. Each trial in each condition comprised 16 quadruplets of words (64 bisyllabic words - 128 syllables - 25.6 seconds), presented consecutively without any silence between them. We used the same syllables in each condition to minimize acoustic variability across conditions. The conditions contain progressively more linguistic information. In Random Syllable Lists, consecutive syllables could not be combined to form real bisyllabic words. This resulted in sequences devoid of linguistic structure at the word, phrase, or sentence levels. In Random Word Lists, two consecutive syllables formed a word, but two consecutive words did not combine into phrases or sentences. However, words within each quadruplet were semantically related and could form a meaningful sentence if reordered correctly. In Fixed Word Lists, consecutive words did not form phrases or sentences, but word categories (e.g., Verb-Noun/Adv-Noun-Adj) were repeated every four words. Words within each quadruplet were not semantically related. Syntactically Anomalous Sentences, are similar to Fixed Word Lists in the sense that word categories were repeated every four words (Verb-Noun/Adv-Noun-Adj). However, the words within each quadruplet were semantically related and could form a meaningful sentence if reordered correctly. In Semantically Anomalous Sentences, quadruplets were structured such that the first two words formed an anomalous noun phrase and the last two an anomalous verb phrase (e.g., Adj-Noun-Noun/Adv-Verb). This created sentences that were syntactically coherent but semantically nonsensical. Last, Control Sentences were composed of quadruplets that formed fully coherent sentences, both syntactically and semantically (Adj-Noun-Noun/Adv-Verb).

We generated 320 bi-syllabic Turkish words, comprising 80 adjectives, nouns, adverbs, and verbs, each intended for use in the condition trials, using the Google Cloud Text-to-Speech service (Google Cloud, 2023), with a female voice (tr-TR-wavenet-A). Additionally, 160 bi-syllabic buffer words were generated for use in the task trials which was to detect the last word presented when the stimuli presentation stop at random time intervals and before condition trials which was excluded from the analysis to due acoustic onset responses.

The sound wave of each bisyllabic word was manually split into two syllables using the software Praat (Boersma and Weenink, 2018). Subsequently, the duration of each syllable was adjusted to 200 ms yielding a syllable rate of $1/0.2s = 5$ Hz, a word rate of $1/0.4 = 2.5$ Hz, a phrase rate of $1/0.8 = 1.25$ Hz and a sentence rate of $1/1.6 = 0.625$ Hz by stretching or compressing the sound wave using a Time-Scale Modification (TSM) algorithm, which alters the length of an audio signal while preserving its pitch (Driedger and Müller, 2014). Additionally, the intensity was normalized to -16 dBFS. In all, this procedure made the words sound more natural than when they were constructed by concatenating separately generated syllables.

To quantify the lexico-semantic relatedness of words in each quadruplet, we calculated the cosine similarity of the word embedding of each word pair in the quadruplets, and then ran a two-sided t-test between the cosine similarity values of semantically related conditions (Random Words, Syntactically Anomalous and Control Sentences) and unrelated conditions (Fixed Word List and Semantically Anomalous Sentences. Word embeddings were obtained from a GPT2 model trained on a Turkish dataset (ytu-ce-cosmos/turkish-gpt2: Kesgin et al., 2024). We used the same GPT2 model to calculate the transitional probability of each word pair, which is the conditional probability of a word given the previous word in each trial of each condition. Then, the surprisal values of each word were computed via the negative log of this probability.

Procedure

Participants listened to the speech streams played through in-ear headphones while their brain activity was measured with MEG while participants were fixating a cross in the middle of the presentation screen. In each condition, there were 10 distinct syllable streams (each 25.6 seconds) each repeated 2 times, resulting in a total of 20 trials for each condition presented in 10 blocks. Four buffer words were presented both before each trial to avoid the evoked response at the stimulus onset. Buffer words were not included in the analysis. Each block started after 30 seconds of silence to allow participants to sit still and prepare for the block. Audio recordings of each trial were played after a random interval ranging from 2 to 3 seconds. To maintain participant engagement throughout the experiment, 2-4 multiple-choice questions were included in each block. These questions were randomly presented at the end of trials, requiring participants to choose the last word they heard from four options by pressing a button. To prevent disruption of condition sequences by the button press during the task, new sets of buffer word sequences, ranging from 10 to 44 words in length, were generated specifically for use during these tasks and they were added at the end of conditions sequences.

Data Acquisition

The magnetoencephalography (MEG) signals were recorded with a 275-sensor axial gradiometer system (CTF Systems Inc.) in a magnetically shielded room. All stimuli were presented auditorily using the Psychophysics Toolbox extensions of Matlab (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). MEG data were acquired at a sampling frequency of 1200 Hz. Head localization was monitored during the experiment using marker coils placed at the cardinal points of the head (nasion, left and right ear canal) and head position was corrected before each block to keep it at the same position as at the beginning of the experiment. In addition to MEG data, we also acquired T1-weighted structural MR images of each participant using a 3-T MAGNETOM Skyra scanner (Siemens Healthcare, Erlangen, Germany). Lastly, 3-dimensional coordinates of each participant's head surface were measured using a digitizing pen system (Polhemus Isotrak system, Kaiser Aerospace Inc.).

MEG Data Preprocessing

MEG data were analyzed using mne-python (version 0.23.1). First, data were annotated to exclude the response parts from the rest of the analysis, resampled to 300 Hz, and then we applied a notch filter at 50 and 100 Hz. Data were epoched into trials from 2 seconds before the onset of the first buffer word to the offset of the last buffer word. Epochs were filtered between 0.5 and 145 Hz with a one-pass, zero-phase, non-causal FIR filter using the default settings of mne-python. Ocular and cardiac artifacts were removed with independent components analysis.

Individual head models were created for each participant with their structural MR images with Freesurfer (surfer.nmr.mgh.harvard.edu) and were co-registered to the MEG coordinate system with mne's coregistration utility. A surface-based source space was computed for each participant using four-fold icosahedral subdivision. Resting state data 2 seconds before the presentation of each trial (4 mins in total) were used to calculate the noise covariance matrix. Cortical sources of the MEG signals were estimated using dynamic statistical parametric map (dSPM), a noise-normalized minimum norm estimate method. Orientations of the dipoles were constrained to be perpendicular to the cortical surface. The resulting source space was morphed to the fsaverage template provided by FreeSurfer. For the subsequent analyses, the initial 1.6 seconds of each epoch, corresponding to the presentation of buffer words, were excluded to avoid confounding effects from the acoustic onset response.

Spectral analysis

Stimuli: The auditory streams for each condition were filtered between 0.5 and 145 Hz using a one-pass, zero-phase, non-causal FIR filter using mne-python library. Then, the temporal envelope was obtained by taking the absolute value of the analytic signal derived via the Hilbert transform. After down-sampling to 50 Hz, the power spectra were computed via a discrete Fourier transform of the envelopes of the stimuli.

Induced Power: The source-reconstructed data were first down-sampled to 50 Hz and then transformed into the frequency domain using a discrete Fourier transform with a frequency resolution of 1/25.6 Hz. The power spectrum for each source point and trial was calculated by taking the magnitude of the complex values, which involves computing the absolute value of the real and imaginary components. Following this, the magnitude of the power across all trials was averaged, and the logarithm of these averaged values was subsequently computed.

The inter-trial phase coherence: ITPC at a given frequency f was computed as the mean of the squared sum of the Fourier coefficients at f. This is calculated using the formula:

$$R(f) = \frac{1}{k}\left[\sum_{k}(\cos\theta_k)^2 + \sum_{k}(\sin\theta_k)^2\right]$$

where k represents the number of trials for each condition, and θ_k denotes the phase angle of the complex Fourier coefficient.

Connectivity Analysis

For the connectivity analysis, the epoched data for each condition and trial were filtered at each frequency of interest (FOIs: 0.625 Hz, 1.25, 2.5, and 5.0 Hz) using a narrow-band, one-pass, zero-phase, non-causal FIR filter with the MNE-Python library. The filter was applied within a range of ± half the spectral resolution around each FOI. Time series of source estimates for each condition and trial were computed for the frontal, parietal, and temporal brain regions using the 'PALS_B12_Brodmann' atlas (Van Essen, 2005), resulting in 24 parcellations per hemisphere. We used pairwise phase consistency (PPC) as the connectivity metric, because PPC is invariant to the direction of information flow, it captures both zero and non-zero phase lag interactions between signals, and it is not biased by the sample size (Vinck et al., 2010, 2012; Bastos & Schoffelen, 2016).

Hierarchically Coupled Oscillatory Network Model

We developed a hierarchically coupled oscillatory network to simulate neural responses across different experimental conditions to test structure-building hypothesis. The network consists of three layers. The first layer includes four oscillators representing word categories (Adjective, Noun, Noun, Verb) for all conditions except the Random Syllable List condition, which does not contain word categories. The second layer contains two phrase nodes. The first phrase node is coupled to the Adjective and Noun oscillators, while the second phrase node is coupled to the Noun and Verb oscillators. In the third layer, a single sentence node is coupled to the phrase nodes from the second layer. The intrinsic frequencies of the oscillators were set to 2.5 Hz for the first layer, 1.25 Hz for the second layer, and 0.625 Hz for the third layer. Linear harmonic oscillators at each layer were modeled as masses connected by springs with a damping factor (Figure 1 C).

The dynamics of each oscillator were governed by the following equation,

$$\ddot{x}(t) = F - \omega^2 x(t) - \gamma \dot{x}(t) \qquad \text{(Eq. 1)}$$

where t is time, x(t) is the magnitude of the neural response, $\dot{x}$ (t) and $\ddot{x}$ (t) are the first and second derivatives of the neural response, respectively, F(t) represents the stimulation input, ω is the intrinsic angular frequency of the oscillator, and γ is the damping factor. The intrinsic angular frequency of the oscillators is determined by the following formula where *f* is the frequency.

$$\omega = 2\pi f \qquad \text{(Eq. 2)}$$

To establish interactions between nodes across layers, a coupling constant C_{ij} was introduced between coupled nodes i and j. The dynamic equations governing all seven nodes are given below.

$$\ddot{x}_1(t) = F_1 - \omega_1^2 x_1(t) + C_{15} x_5(t) - \gamma_1 \dot{x}_1(t) \qquad \text{(Eq. 3)}$$

$$\ddot{x}_2(t) = F_2 - \omega_2^2 x_2(t) + C_{25} x_5(t) - \gamma_2 \dot{x}_2(t)$$

$$\ddot{x}_3(t) = F_3 - \omega_3^2 x_3(t) + C_{36} x_6(t) - \gamma_3 \dot{x}_3(t)$$

$$\ddot{x}_4(t) = F_4 - \omega_4^2 x_4(t) + C_{46} x_6(t) - \gamma_4 \dot{x}_4(t)$$

$$\ddot{x}_5(t) = -\omega_5^2 x_5(t) + C_{15} x_1(t) + C_{25} x_2(t) + C_{57} x_7(t) - \gamma_5 \dot{x}_5(t)$$

$$\ddot{x}_6(t) = -\omega_6^2 x_6(t) + C_{36} x_3(t) + C_{46} x_4(t) + C_{67} x_7(t) - \gamma_6 \dot{x}_6(t)$$

$$\ddot{x}_7(\mathrm{t}) = -\omega_7^2 x_7(\mathrm{t}) + C_{57}x_5(\mathrm{t}) + C_{67}x_6(\mathrm{t}) - \gamma_7\dot{x}_7(t)$$

To simplify the model and have a symmetrical network to ensure the stability of the network, we set the forward and backward coupling strengths to be equal. Both damping factors and coupling strengths were scaled proportionally to the oscillator constant (ω^2). The oscillator constants for oscillators 1 through 4 were set to 246.74, corresponding to a natural frequency of 2.5 Hz as defined by Equation 2. For oscillators 5 and 6, the constants were set to 61.68, yielding a frequency of 1.25 Hz, and for oscillator 7, the constant was 15.42, corresponding to a frequency of 0.625 Hz.

Effect of damping factor:

The damping factor (γ) for word-level oscillators was set to 8 to ensure that the amplitude of oscillatory activity remained stable throughout the stimulation period in the absence of coupling with slower oscillators (Figure 7). At lower damping values, external stimulation leads to an increase in the amplitude of total oscillatory activity, potentially confounding the effects of coupling. For higher-level oscillators, the damping factors were reduced in proportion to their respective oscillator constants, since slower oscillators decay more quickly when using the same damping value. Specifically, the damping factor was set to 2 for phrase-level oscillators and 0.5 for sentence-level oscillators.

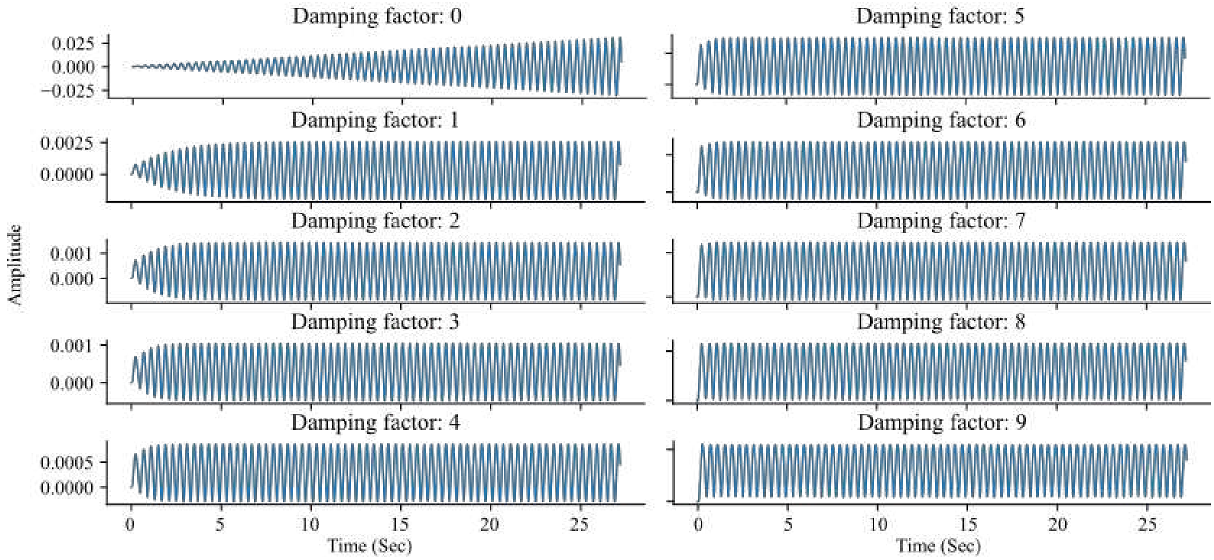


Figure 7. Amplitude of total activity in Word Level oscillators with varying damping factors

Effect of coupling constants:

The coupling constants were set to be smaller than the oscillator's intrinsic constant ω^2 to ensure that bottom-up stimulation does not dominate the dynamics of higher-level oscillators, maintaining a weakly coupled regime. Additionally, the coupling strength of the second lower-level node to the higher-level node was chosen to be smaller (%20) than that of the first node. This reduces the suppression caused by the down-phase of the slower oscillation from the higher level, allowing the higher-level oscillator to produce a stronger peak when integrating input from both lower-level nodes. Functionally, this may reflect reduced inhibition for competing words in the second word category, following syntactic predictions initiated by the first word category (see section Effect of coupling strength asymmetry).

Word nodes were stimulated using Gaussian noise with a mean of 1 and 10% standard deviation, applied over a 400 ms window. To smooth the onset and offset of the stimulation, a linear fade-in and fade-out of 80 ms was applied (Figure 8). Each word category node was stimulated in the same sequence as in the experimental conditions. Consistent with the experimental design, we presented 17-word quadruplets and excluded the first quadruplet from the analysis. The total simulation duration was 25.6 seconds, matching the previous experiment. The model was simulated 30 times with varying input signals by numerically solving the differential equations using the solve_ivp function from the SciPy library for Python (Virtanen et al., 2020). All oscillators were initialized with zero initial values. Then, a Fourier transform was applied to the aggregated activation of all oscillators.

We began our simulations using the Control Sentences condition, which exhibited the highest amplitude at sentence and phrase level frequencies in the experiment, to evaluate the impact of coupling strengths between network nodes. Since the Syntactically Anomalous condition involved a reversed presentation of the Control Sentences, and the Random Word List condition consisted of the same words presented in a randomized order within each quadruplet, we simulated these two conditions by keeping all network parameters constant and varying only the stimulation sequence, in alignment with the design of the MEG experiment.

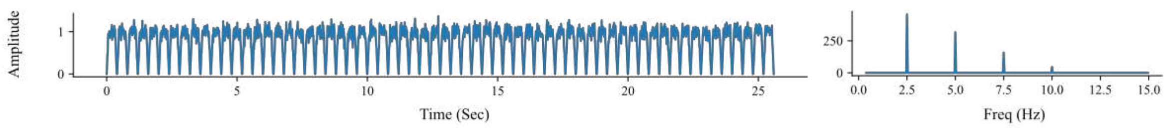


Figure 8 A) Input stimulating word nodes **B)** Power spectrum of the input

Effect of coupling strength asymmetry:

We then examined the effects of coupling strength asymmetry between the first and second word-category nodes by systematically varying the ratio of their coupling strengths. The results indicated that when the coupling strengths were equal (ratio = 1), there was no prominent power peak at the phrase- and sentence-level frequencies. However, as the ratio deviated from 1, power at these slower frequencies increased (Figure 9 A, B). This effect arises because, during the down-phase of the slower oscillation, initiated by the first node of the faster oscillator, the second node becomes active and reinforces the upward phase, thereby driving the slower oscillator. When the coupling strengths are equal, the slower oscillator is entrained by the faster one and begins oscillating at the same frequency, suppressing the natural frequency of slow oscillator. In contrast, when the coupling strength of the second node is weaker than the first node, the slower oscillator can resonate at its intrinsic frequency, resulting in a distinct spectral peak (Figure 9 D).

When the coupling strength of the second node is stronger than the first node, a peak at slower frequencies still emerges. However, in this case, the amplitude is greater for the reverse-order condition (*Syntactically Anomalous*) compared to the normal-order condition (*Control Sentences*), which contradicts empirical observations. Furthermore, amplitude differences across the three conditions, normal (*Control Sentences*), reverse (*Syntactically Anomalous*), and random (*Random Word List*), were most balanced when the coupling strength ratio was set to 0.8. As the ratio deviated further from 1, the difference between the normal and reverse conditions diminished compared to between reverse and random, despite the increase in power at the phrase and sentence frequencies (Figure 4 A and C). Based on these observations, we selected a coupling strength ratio of 0.8 to better capture the condition-specific differences in amplitude dynamics.

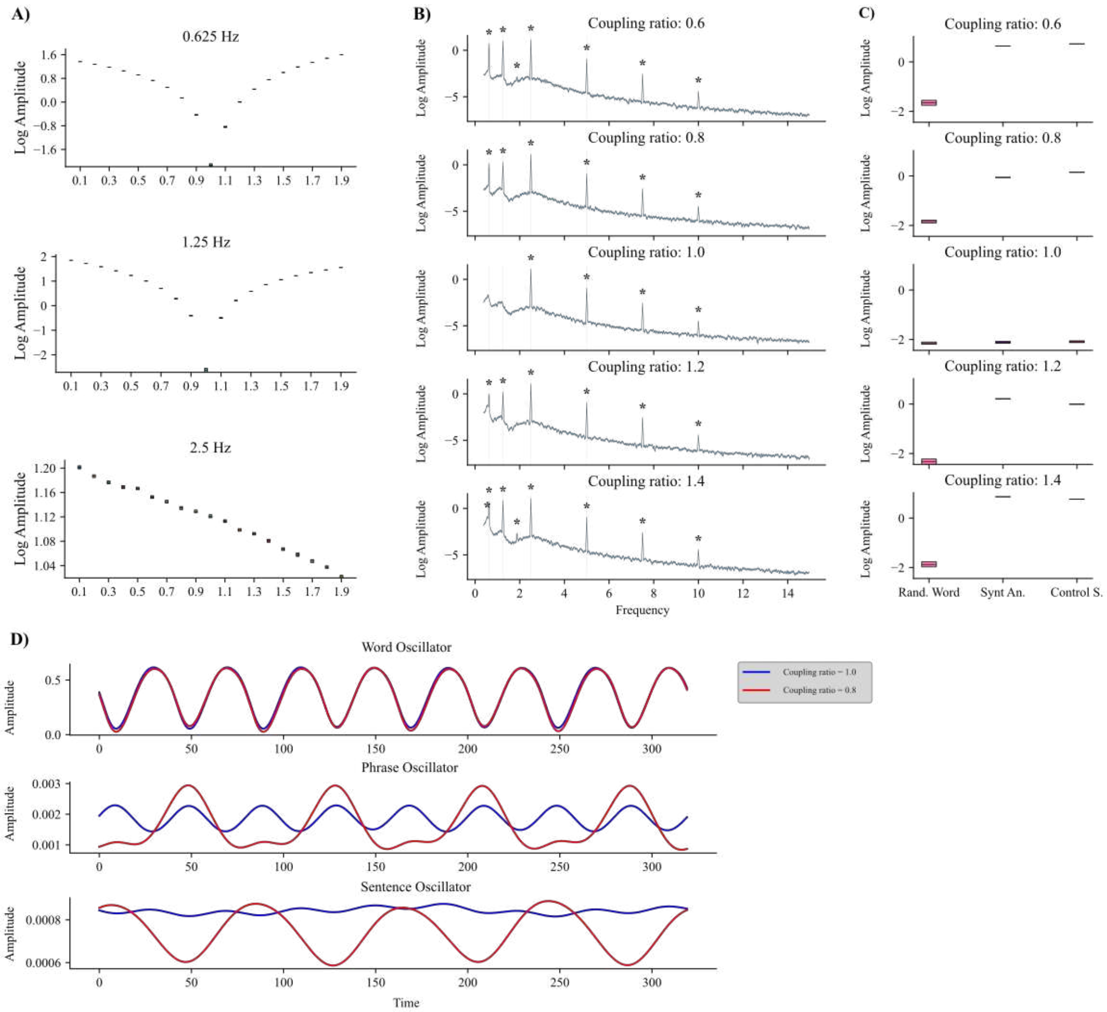


Figure 9. A) Changes in log power at sentence-, phrase-, and word-level frequencies as a function of the coupling strength ratio between the first and second daughter nodes. **B)** Power spectra across varying coupling strength ratios. (* < 0.05) **C)** Sentence-level power for the Random Words, Syntactically Anomalous, and Control Sentences conditions across different coupling ratios. **D)** Amplitudes of word, phrase and sentence level nodes at coupling ratios of 0.8 and 1.0. (Vertical lines show the onset of stimulus)

Effect of coupling strengths:

We then simulated the network for *Control Sentences* conditions using a range of coupling strengths between the first and second layers, and between the second and third layers, to examine the effects of coupling on network dynamics. Coupling strengths between the first and second layers were varied from 2 to 30, ensuring that the combined input from the two nodes did not exceed the intrinsic oscillator constant ω^2 of nodes 5 and 6 (61.68). Similarly, coupling strengths between the second and third layers were varied from 1 to 7, also to keep the total input within the bounds set by the ω^2 values of the receiving nodes. Increasing the coupling strength was found to enhance the power of sentence- and phrase-level frequencies while simultaneously reducing power at the word-level frequency (Figure 10). To replicate patterns seen in the experimental data, we selected a coupling strength of 28 between sentence and phrase frequencies to optimize amplitudes at sentence and phrase level peaks, and 7 for the coupling from phrase to sentence nodes values scaled in proportion to their respective oscillator constants.

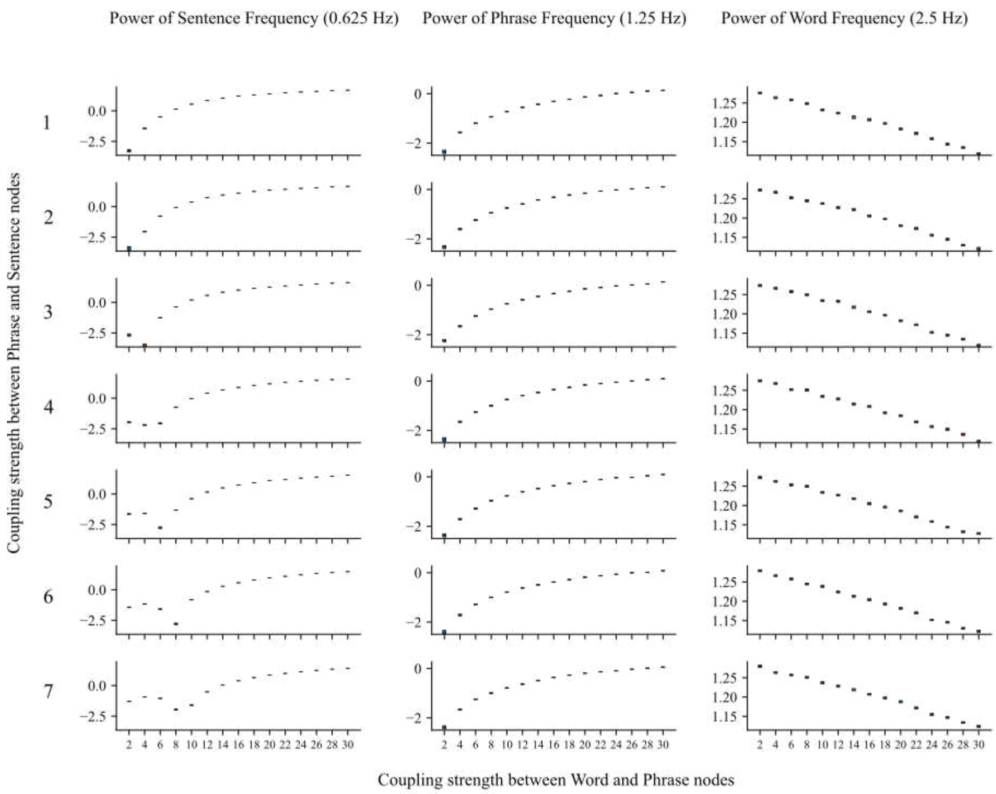


Figure 10. Log power at sentence-, phrase-, and word-level frequencies as a function of coupling strength between word and phrase nodes. The x-axis in each graph represents the

coupling strength between word and phrase nodes, while each row corresponds to a different fixed coupling strength between phrase and sentence nodes.

To simulate the *Semantically Anomalous* and *Fixed Word Order* conditions, characterized by reduced transitional probability (and/or semantic relatedness among words) within each quadruplet, we maintained the same connectivity structure between word, phrase, and sentence nodes. However, based on prior simulation results indicating that both weaker inter-layer coupling and more symmetric intra-layer coupling reduce power at phrase- and sentence-level frequencies, we adjusted the coupling parameters accordingly. Specifically, we assumed that the coupling constants between hierarchical layers should be reduced and that the coupling strength ratio between the first and second daughter nodes should be closer to 1.

In the model, the network's connectivity pattern represents syntactic structure, while coupling strength encodes statistical relationships among nodes. Experimental findings from the MEG study showed that phrase- and sentence-level power in the *Semantically Anomalous* condition was lower than in the *Control Sentences* condition but higher than in the *Syntactically Anomalous* condition. Moreover, there was no significant difference in power between the *Fixed Word Order* condition, which is a reversed presentation of the *Semantically Anomalous* condition, and the *Syntactically Anomalous* condition.

To replicate these empirical results, we gradually reduced the inter-layer coupling strengths and adjusted the daughter-node coupling ratio. A close match to the experimental data was achieved when the coupling strength ratio between daughter nodes was set to 0.82 (as opposed to 0.8 in the baseline model for Control Sentences) and the overall coupling strengths were reduced by 5%. Similar to experimental results, all conditions exhibited a significant peak ($p < 0.05$) at the word (2.5 Hz) and syllable (5 Hz) frequencies. The Control Sentences, Semantically Anomalous Sentences, Syntactically Anomalous Sentences, and

Fixed Word List conditions additionally showed significant peaks at both sentence (0.625 Hz) and phrase (1.25 Hz) frequencies. (Figure 5 A).

Lexical Model

Using an approach similar to Frank and Yang's (2018) lexical model to test the lexical-statistics hypothesis (Burroughs et al., 2021; Lo et al., 2022; Kalenkovich et al., 2022) between conditions, we simulated a stimulus time series in the following way. 768-dimensional word-embedding vectors from the GPT-2 model were repeated across 40 columns to simulate a word duration of 400 ms at a sampling rate of 100 Hz. The onset time *t* for each word was randomly drawn from a uniform distribution with a mean of 40 ms and a width of 50 ms. These word embeddings were concatenated into sequences of 16 four-word sentences to represent individual trials. Gaussian noise with a standard deviation of 0.05 was added to the trial matrices, and the discrete Fourier transform was applied to each row. The resulting spectral power was averaged across rows to produce a single power spectrum for each trial in each condition. The Random Syllable List condition was not included because it does not contain real words and therefore could not be modeled via word embeddings.

Statistical analysis

To compare the power spectrum of stimuli, simulated response with lexical model at each frequency of interest, and the surprisal values of words in trials between conditions, we run an ANOVO test after checking for normality with Shapiro-Wilk test. When we found a significant main effect for condition, we carried out a posteriori analysis for pairwise comparison using Holm procedure to correct for multiple comparison using scikit toolbox in Python.

For both power and ITPC, we tested if the ITPC and power at each frequency bin was higher than the average of two neighboring frequency bins on each side by using non-parametric 1-

sided permutation tests with 10.000 permutations (Maris and Oostenveld, 2007) This analysis was done to detect significant peaks at the different frequencies of interest.

To compare the spectral power and ITPC across conditions, we fitted a linear mixed effect model using the lmer function in the lme4 package for R (R Core Team, 2024) with Condition as fixed effect and subjects as random effect. The dependent variable was the power and ITPC averaged over all source points. Separate models were run for each frequency of interest (FOIs: sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz), and syllable (5.0 Hz)). We used the "emmeans" function (Lenth et al., 2018) in R for posthoc pairwise comparisons.

We compared the connectivity difference between conditions on the brain parcellations for each frequency bin using cluster-based permutation test (Maris & Oostenveld, 2007) with 10,000 permutations.

Supplementary Materials

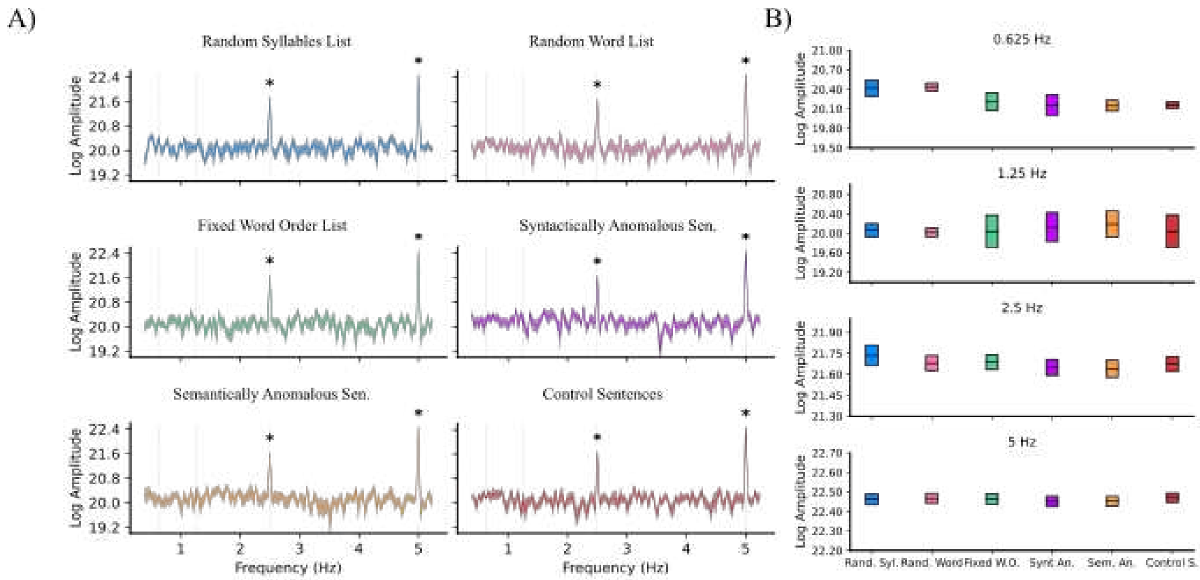


Figure 1. A) Power spectrum of stimuli in each condition. The vertical dashed lines indicate the FOIs at the sentence (0.625 Hz), phrase (1.25 Hz), word (2.5 Hz), and syllable frequency

(5 Hz), and black asterisks indicate significant peaks at FOIs, and the grey asterisks indicate significant peaks at the harmonic frequencies. Shaded areas show the standard error of the mean within trials **B)** Comparison of power averaged over trials. Box edges indicate the standard error of the mean over trials. Black line shows the mean.

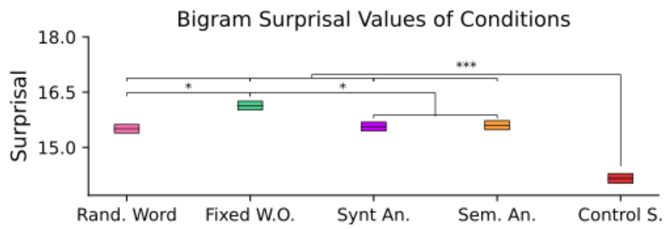


Figure 2. Comparison of surprisal values of each word in conditions. Box edges indicate the standard error of the mean over trials. Black line shows the mean.

Table 1. Adjusted p values of pairwise comparison of surprisal values between conditions

Comparison	p adj
Random Words - Fixed Word Order	0.0009
Random Words - Syntactically Anomalous	1
Random Words - Semantically Anomalous	1
Random Words - Control Sentences	<.0001
Fixed Word Order - Syntactically Anomalous	0.0045
Fixed Word Order - Semantically Anomalous	0.0077
Fixed Word Order - Control Sentences	<.0001
Syntactically Anomalous - Semantically Anomalous	1
Syntactically Anomalous - Control Sentences	<.0001
Semantically Anomalous - Control Sentences	<.0001

Table 2. Adjusted p values of pairwise comparison of induced power of the neural signal between conditions

Comparison	p adj - 0.625 Hz	p adj - 1.25 Hz	p adj - 2.5 Hz	p adj - 5 Hz
Random Syllables - Random Words	0.9991	0.9996	<.0001	0.0001
Random Syllables - Fixed Word Order	0.241	0.0185	<.0001	0.0237
Random Syllables - Syntactically Anomalous	0.9466	0.0094	<.0001	0.0146
Random Syllables - Semantically Anomalous	0.0917	<.0001	<.0001	0.008
Random Syllables - Control Sentences	0.0001	<.0001	<.0001	0.0026
Random Words - Fixed Word Order	0.4453	0.046	0.9996	0.6807

Random Words - Syntactically Anomalous	0.995	0.0248	1	0.776
Random Words - Semantically Anomalous	0.2082	<.0001	0.9837	0.869
Control Sentences - Random Words	0.0003	<.0001	0.1811	0.9665
Fixed Word Order - Syntactically Anomalous	0.7794	0.9999	0.9998	1
Fixed Word Order - Semantically Anomalous	0.9978	0.0248	0.9217	0.9993
Fixed Word Order - Control Sentences	0.1171	<.0001	0.0911	0.9848
Syntactically Anomalous - Semantically Anomalous	0.4985	0.0459	0.9799	1
Syntactically Anomalous - Control Sentences	0.0023	<.0001	0.1691	0.9956
Semantically Anomalous - Control Sentences	0.2911	0.1844	0.555	0.9995

Table 3. Adjusted p values of pairwise comparison of ITPC of the neural signal between conditions

Comparison	**p adj - 0.625 Hz**	**p adj - 1.25 Hz**	**p adj - 2.5 Hz**	**p adj - 5 Hz**
Random Syllables - Random Words	0.5227	0.9999	<.0001	0.0035
Random Syllables - Fixed Word Order	<.0001	<.0001	<.0001	0.0006
Random Syllables - Syntactically Anomalous	<.0001	<.0001	<.0001	0.0009
Random Syllables - Semantically Anomalous	<.0001	<.0001	<.0001	0.0701
Random Syllables - Control Sentences	<.0001	<.0001	<.0001	<.0001
Random Words - Fixed Word Order	<.0001	<.0001	1	0.9966
Random Words - Syntactically Anomalous	<.0001	<.0001	0.991	0.9991
Random Words - Semantically Anomalous	<.0001	<.0001	0.8691	0.9229
Random Words - Control Sentences	<.0001	<.0001	0.1568	0.8168
Fixed Word Order - Syntactically Anomalous	0.8217	1	0.9759	1
Fixed Word Order - Semantically Anomalous	0.3895	<.0001	0.7983	0.685
Fixed Word Order - Control Sentences	<.0001	<.0001	0.1127	0.9746
Syntactically Anomalous - Semantically Anomalous	0.0233	<.0001	0.9949	0.7539
Syntactically Anomalous - Control Sentences	<.0001	<.0001	0.4561	0.9537
Semantically Anomalous - Control Sentences	<.0001	0.0346	0.7898	0.2343

Table 4. Adjusted p values of pairwise comparison of power of the lexical model between conditions

Comparison	**p adj - 0.625 Hz**	**p adj - 1.25 Hz**	**p adj - 2.5 Hz**	**p adj - 5 Hz**
Random Words - Fixed Word Order	<.0001	1	1	1
Random Words - Syntactically Anomalous	<.0001	1	1	1
Random Words - Semantically Anomalous	<.0001	1	1	1
Random Words - Control Sentences	<.0001	1	1	1
Fixed Word Order - Syntactically Anomalous	1	1	1	1
Fixed Word Order - Semantically Anomalous	1	1	1	1

Fixed Word Order - Control Sentences	1	1	1	1
Syntactically Anomalous - Semantically Anomalous	1	1	1	1
Syntactically Anomalous - Control Sentences	1	1	1	1
Semantically Anomalous - Control Sentences	1	1	1	1

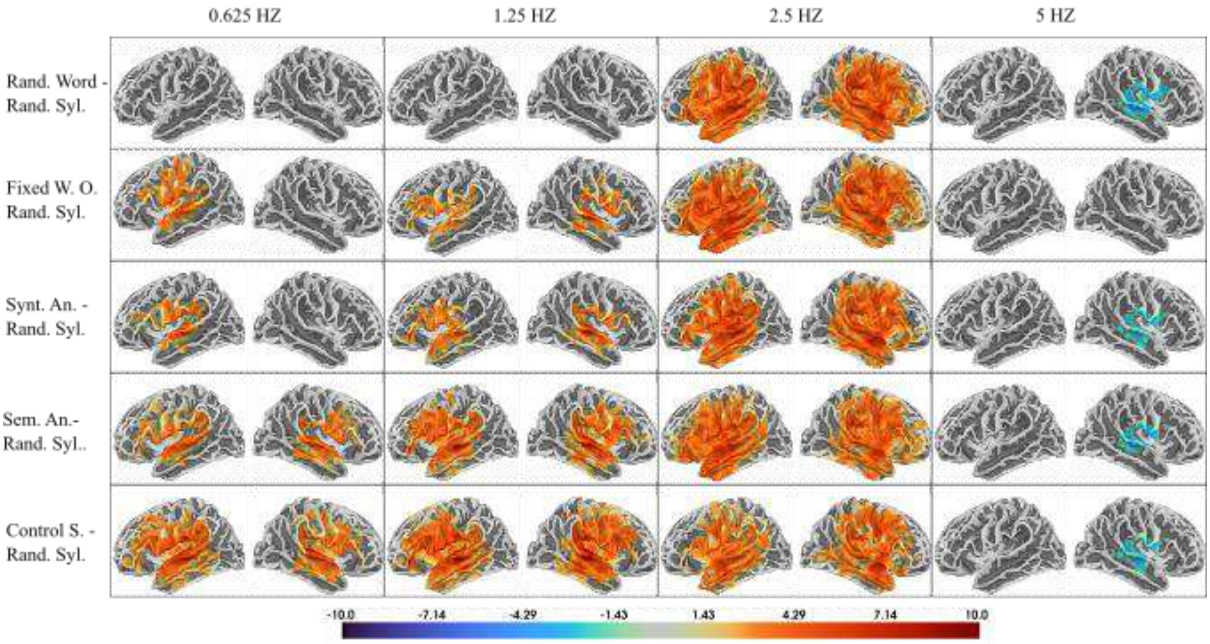


Figure 3. Sources that were significantly different between the contrasts. The colors indicate the t-values of the corresponding comparisons.

ACKNOWLEDGEMENTS

AEM was supported by an Independent Max Planck Research Group and a Lise Meitner Research Group “Language and Computation in Neural Systems” from the Max Planck Society, by the Netherlands Organization for Scientific Research (NWO; VIDI grant 016.Vidi.188.029 and Aspasia grant 015.014.013).

COMPETING INTERESTS: No competing interests declared.

REFERENCES

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, *25*(9), 3077-3085.

Bai, F., Meyer, A. S., & Martin, A. E. (2022). Neural dynamics differentially encode phrases and sentences during spoken language comprehension. *PLoS Biology*, *20*(7), e3001713.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. Cortex, 115, 56-71.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: a (re-) emerging program of research for language and vision. *Biolinguistics*, *4*(2-3), 174-200.

Brennan, J. R., & Martin, A. E. (2020). Phase synchronization varies systematically with linguistic structure composition. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20190305.

Brodbeck, C., Kandylaki, K. D., & Scharenborg, O. (2024). Neural representations of non-native speech reflect proficiency and interference from native language knowledge. Journal of Neuroscience, 44(1).

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. Current Biology, 28(5), 803-809.

Burroughs, A., Kazanina, N., & Houghton, C. (2021). Grammatical category and the neural processing of phrases. *Scientific reports*, *11*(1), 2446.

Coopmans, C. W., De Hoop, H., Hagoort, P., & Martin, A. E. (2022). Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. Neurobiology of Language, 3(3), 386-412.

Coopmans, C. W., De Hoop, H., Tezcan, F., Hagoort, P., & Martin, A. E. (2025). Language-specific neural dynamics extend syntax into the time domain. PLoS biology, 23(1), e3002968.

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences*, *90*(11), 4957-4960.

Di Liberto, G., O'Sullivan, J., & Lalor, E. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology, 25(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

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

Doelling, K. B., & Assaneo, M. F. (2021). Neural oscillations in speech and language processing. Nature Reviews Neuroscience, 22(6), 371–384.

Donhauser, P. W., & Baillet, S. (2020). Two distinct neural timescales for predictive speech processing. Neuron, 105(2), 385-393.

Frank, S. L., & Christiansen, M. H. (2018). Hierarchical structure in language and the brain. Cognition, 176, 59–70.

Frank, S. L., & Yang, J. (2019). Neural tracking of linguistic structure: A computational perspective. Language, Cognition and Neuroscience, 34(9), 1135–1145.

Frank, S. L., & Yang, J. (2019). Neural tracking of linguistic structure: A computational perspective. Language, Cognition and Neuroscience, 34(9), 1135–1145.

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use?. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1747), 4522-4531.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. Trends in cognitive sciences, 6(2), 78-84.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function.

Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B, 360(1456), 815–836.

Ghitza, O. (2020). "Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019). *Language, Cognition and Neuroscience*, *35*(9), 1100-1105.

Gillis, M., Vanthornhout, J., & Francart, T. (2023). Heard or understood? Neural tracking of language features in a comprehensible story, an incomprehensible story and a word list. eneuro, 10(7).

Giraud, A. L. (2020). Oscillations for all_ (ツ) _/? A commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, *35*(9), 1106-1113.

Gwilliams, L. (2020). Hierarchical oscillators in speech comprehension: A commentary on Meyer, Sun, and Martin (2019). *Language, Cognition and Neuroscience*, *35*(9), 1114-1118.

Haegens, S. (2020). Entrainment revisited: a commentary on Meyer, Sun, and Martin (2020). *Language, Cognition and Neuroscience*, *35*(9), 1119-1123.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. Trends in cognitive sciences, 9(9), 416-423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00416

Halle, M., & Stevens, K. N. (1962). Speech Recognition: A Model and a Program. *1962.*

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. Proceedings of the National Academy of Sciences, 119(32), e2201968119.

Holmes, J. M., Marshall, J. C., & Newcombe, F. (1971). Syntactic class as a determinant of word-retrieval in normal and dyslexic subjects. *Nature*, *234*(5329), 418-418.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A. L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *elife*, *4*, e06213.

Izhikevich, E. M., Desai, N. S., Walcott, E. C., & Hoppensteadt, F. C. (2003). Bursts as a unit of neural information: selective communication via resonance. Trends in neurosciences, 26(3), 161-167.

Kalenkovich, E., Shestakova, A., & Kazanina, N. (2022). Frequency tagging of syntactic structure or lexical properties; a registered MEG study. *Cortex*, *146*, 24-38.

Kandylaki, K. D., & Kotz, S. A. (2020). Distinct cortical rhythms in speech and language processing and some more: a commentary on Meyer, Sun, & Martin (2019). *Language, Cognition and Neuroscience*, *35*(9), 1124-1128.

Karunathilake, I. D., Brodbeck, C., Bhattasali, S., Resnik, P., & Simon, J. Z. (2025). Neural dynamics of the processing of speech features: Evidence for a progression of features from acoustic to sentential processing. *Journal of Neuroscience*, *45*(11).

Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. Journal of Neuroscience, 40(49), 9467-9475.

Klimovich-Gray, A., & Molinaro, N. (2020). Synchronising internal and external information: a commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, *35*(9), 1129-1132.

Kösem, A., Basirat, A., Azizi, L., & van Wassenhove, V. (2018). High-frequency neural activity predicts word parsing in ambiguous speech streams. Journal of Neuroscience, 38(29), 7189–7199.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. science, 320(5872), 110-113.

Lewis, A. G. (2020). Balancing exogenous and endogenous cortical rhythms for speech and language requires a lot of entraining: a commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, *35*(9), 1133-1137.

Lisman, J. E., & Jensen, O. (2013). The theta gamma neural code. Neuron, 77(6), 1002–1016.

Lo, C. W., Tung, T. Y., Ke, A. H., & Brennan, J. R. (2022). Hierarchy, not lexical regularity, modulates low-frequency neural synchrony during language comprehension. *Neurobiology of Language*, *3*(4), 538-555.

Lu, Y., Jin, P., Pan, X., & Ding, N. (2022). Delta-band neural activity primarily tracks sentences instead of semantic properties of words. *NeuroImage*, *251*, 118979.

Mai, G., & Wang, W. S. Y. (2023). Distinct roles of delta-and theta-band neural tracking for sharpening and predictive coding of multi-level speech features during spoken language processing. Human Brain Mapping, 44(17), 6149-6172.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29-63.

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. Frontiers in psychology, 7, 120.

Martin, A. E. (2020). A compositional neural architecture for language. Journal of Cognitive Neuroscience, 32(8), 1407-1427.

Martin, A. E., & Doumas, L. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. PLoS Biology, 15(3), e2000663.

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Language, Cognition and Neuroscience, 35(9), 1089–1099.

Murphy, E. (2024). ROSE: A neurocomputational architecture for syntax. Journal of Neurolinguistics, 70, 101180.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649-1653.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A. L. (2017). θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech.

Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and cognitive processes*, *26*(7), 935-951.

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*(6461), 62-66.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature reviews neuroscience*, *18*(1), 42-55.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79–87. https://doi.org/10.1038/4580

Rogers, T. T., Cox, C. R., Lu, Q., Shimotake, A., Kikuchi, T., Kunieda, T., ... & Ralph, M. A. L. (2021). Evidence for a deep, distributed and dynamic code for animacy in human ventral anterior temporal cortex. *elife*, *10*, e66276.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926–1928.

Schoffelen, J. M., Hultén, A., Lam, N., Marquand, A. F., Uddén, J., & Hagoort, P. (2017). Frequency-specific directed interactions in the human brain network for language. *Proceedings of the National Academy of Sciences*, *114*(30), 8083-8088.

Shapiro, K. A., Mottaghy, F. M., Schiller, N. O., Poeppel, T. D., Flüß, M. O., Müller, H. W., ... & Krause, B. J. (2005). Dissociating neural correlates for nouns and verbs. *Neuroimage*, *24*(4), 1058-1067.

Slaats, S., & Martin, A. E. (2025). What’s surprising about surprisal. *Computational Brain & Behavior*, *8*(2), 233-248.

Slaats, S., Meyer, A. S., & Martin, A. E. (2024). Lexical surprisal shapes the time course of syntactic structure building. Neurobiology of Language, 5(4), 942-980.

Slaats, S., Weissbart, H., Schoffelen, J. M., Meyer, A. S., & Martin, A. E. (2023). Delta-band neural responses to individual words are modulated by sentence processing. Journal of Neuroscience.

Sohoglu, E., Beckers, L., & Davis, M. H. (2024). Convergent neural signatures of speech prediction error are a biological marker for spoken word recognition. *Nature Communications*, *15*(1), 9984.

Ten Oever, S., & Martin, A. E. (2021). An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. Elife, 10, e68066.

Ten Oever, S., Carta, S., Kaufeld, G., & Martin, A. E. (2022). Neural tracking of phrases in spoken language comprehension is automatic and task-dependent. Elife, 11, e77468.

Tezcan, F., Ten Oever, S., Bai, F., te Rietmolen, N., & Martin, A. (2025). Linguistic structure and language familiarity sharpen phoneme encoding in the brain.

Tezcan, F., Weissbart, H., & Martin, A. E. (2023). A tradeoff between acoustic and linguistic feature encoding in spoken language comprehension. Elife, 12, e82386.

Vinck, M., Uran, C., Spyropoulos, G., Onorato, I., Broggini, A. C., Schneider, M., & Canales-Johnson, A. (2023). Principles of large-scale neural interactions. Neuron, 111(7), 987-1002.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272.

Weissbart, H., & Martin, A. E. (2024). The structure and statistics of language jointly shape cross-frequency neural dynamics during spoken language comprehension. *Nature communications*, *15*(1), 8850.

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. Journal of cognitive neuroscience, 32(1), 155-166.

Zhao, J., Martin, A. E., & Coopmans, C. W. (2024). Structural and sequential regularities modulate phrase-rate neural tracking. Scientific Reports, 14(1), 16603.

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. Frontiers in human neuroscience, 10, 274.